

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.24485 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-20
Perfect score: 1248
Sequence: 1 MILEDCGNSLIKRWIEGA.....GARIMPLVFLVGLALACPTE 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	7.1	4150	US-09-428-517-2	Sequence 2, Appli
2	80.5	6.5	5087	US-09-144-085-1	Sequence 1, Appli
3	79	6.3	1832	US-09-335-409-4	Sequence 4, Appli
4	79	6.3	1832	US-09-568-102-4	Sequence 4, Appli
5	79	6.3	1832	US-09-567-969-4	Sequence 4, Appli
6	79	6.3	1832	US-09-568-480-4	Sequence 4, Appli
7	79	6.3	1832	US-09-568-486-4	Sequence 4, Appli
8	79	6.3	1832	US-09-568-472-4	Sequence 4, Appli
9	79	6.3	1832	US-09-567-899-4	Sequence 4, Appli
10	79	6.3	2259	US-09-413-814-70	Sequence 4, Appli
11	79	6.3	2439	US-09-335-409-7	Sequence 70, Appl
12	79	6.3	2439	US-09-568-102-7	Sequence 7, Appli
13	79	6.3	2439	US-09-567-969-7	Sequence 7, Appli
14	79	6.3	2439	US-09-568-480-7	Sequence 7, Appli
15	79	6.3	2439	US-09-568-486-7	Sequence 7, Appli
16	79	6.3	2439	US-09-568-472-7	Sequence 7, Appli
17	79	6.3	2439	US-09-567-899-7	Sequence 7, Appli
18	79	6.3	3567	US-07-642-734C-4	Sequence 4, Appli
19	79	6.3	3567	US-08-439-009A-4	Sequence 4, Appli
20	79	6.3	4551	US-09-320-878-1	Sequence 1, Appli
21	79	6.3	4613	US-09-105-537-31	Sequence 31, Appl
22	79	6.3	11877	US-09-105-537-6	Sequence 6, Appli
23	77	6.2	855	US-08-938-365-3	Sequence 3, Appli
24	77	6.2	867	US-08-938-365-2	Sequence 2, Appli
25	77	6.2	954	US-08-749-169A-3	Sequence 3, Appli
26	77	6.2	954	US-09-130-032A-3	Sequence 3, Appli
27	77	6.2	6095	US-09-144-085-2	Sequence 2, Appli

28	76.5	6.1	1147	1	US-08-144-121-3	Sequence 3, Appli
29	76.5	6.1	1147	2	US-08-735-893-3	Sequence 3, Appli
30	76.5	6.1	1165	1	US-08-144-121-2	Sequence 2, Appli
31	76.5	6.1	1165	2	US-08-735-893-2	Sequence 2, Appli
32	76	6.1	1861	2	US-08-790-912-4	Sequence 4, Appli
33	76	6.1	5588	4	US-09-036-987A-6	Sequence 6, Appli
34	76	6.1	5588	4	US-09-370-700-6	Sequence 6, Appli
35	75	6.0	176	4	US-08-969-683A-65	Sequence 65, Appl
36	75	6.0	1881	4	US-09-233-086-3	Sequence 3, Appli
37	74.5	6.0	1562	3	US-09-320-878-3	Sequence 3, Appli
38	74.5	6.0	1562	4	US-09-105-537-35	Sequence 35, Appl
39	74	5.9	252	4	US-09-199-637A-176	Sequence 176, App
40	74	5.9	493	4	US-09-177-349-5	Sequence 5, Appli
41	73.5	5.9	257	4	US-09-287-097-2	Sequence 2, Appli
42	73.5	5.9	2588	3	US-08-936-135-2	Sequence 2, Appli
43	73	5.8	858	2	US-08-907-166-2	Sequence 2, Appli
44	72.5	5.8	902	4	US-09-134-001C-5157	Sequence 5157, Ap
45	72	5.8	858	3	US-08-946-026-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-428-517-2
; Sequence 2, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; EARLIER FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-428-517-2

Query Match	7.1%	Score 89;	DB 4;	Length 4150;
Best Local Similarity	25.5%	Pred. No. 1.1;	74;	Indels 58; Gaps 9;
Matches	51;	Conservative	17;	Mismatches
QY	3	LEIDCGNSLIKRWIEGAARSVAGGLAESDQALVEQITSOQALFVACRLVSVRSEQETS 62		
Db	2111	LDEDAGRLV-----DVVGLADEQLAV-----RASSVLARLRVTPGHRMSS 2154		
QY	63	QLVARLEQFPVSALVASSGKQLAGVRNGYLDYQRLGLDRLWLALVAHHL-----112		
Db	2155	QAGGR--EWSPSGTVLTGTGTALGAH-----VARWLAGKGAHLVLSRRGADA 2202		
QY	113	AKKACL---VIDGTATVSDVAADGVHLGGYICPGMTLMRSOLRTHTRIRYDDAEAR- 168		
Db	2203	AGAAALRDSLTDMGVRVT--LAACDAAD-----RHALETLLDSLRTPAQLTA 2248		
QY	169	--RALASLPQGQATAEAVR 186		
Db	2249	VIHAGALDDGMTVTLTPEQ 2268		

RESULT 2

US-09-144-085-1
 ; Sequence 1, Application US/09144085
 ; Patent No. 6280999
 ; GENERAL INFORMATION:
 ; APPLICANT: Gustafsson, Claes
 ; APPLICANT: Betlach, Mary C.
 ; APPLICANT: Ashley, Gary
 ; APPLICANT: Julien, Bryan
 ; APPLICANT: Ziermann, Rainer
 ; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
 ; FILE REFERENCE: 30062-20020.20
 ; CURRENT APPLICATION NUMBER: US/09/144,085
 ; EARLIER FILING DATE: 1998-08-31
 ; EARLIER APPLICATION NUMBER: 09/010,809
 ; EARLIER FILING DATE: 1998-01-22
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 5087
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-144-085-1

Query Match 6.5%; Score 80.5; DB 4; Length 5087;
 Best Local Similarity 27.6%; Pred. No. 17;
 Matches 56; Conservative 24; Mismatches 78; Indels 45; Gaps 10;
 QY 26 GGLAESDVALEQL-----TSQALPVRACRLVSRSEQETSOLVARLEQLFPVS 75
 DB 1224 GAVSAAPDVAEDLAHGLWGLIRTARSEHFERILIDVTEPVDAGLLAR----- 1275
 QY 76 ALVASSGKQLGVRNGYLDYQRLGDRWLALVAHHAKKACL-----VIDLGT 124
 DB 1276 ALATAAEPELA-LGGAVLAARL---VRVQAAAEELTRAGLDPACTVLVTGAVGLGQ 1330
 QY 125 AVTSDVAADGV-HL-----GGYICPG-WTLMRS--QLRTHTRIRYDDAEARRALASLQ 175
 DB 1331 AVTRHVRHAGVRLVLTLSRGLGAPGARELVQSLEELGAEVTSWVACDVSKREEIARVL 1390
 QY 176 PGQATAEAVRGCLLMRLRGFVRE 198
 DB 1391 AGIDAARPL--SAVLHLAGVVHD 1411

RESULT 3
 US-09-335-409-4
 ; Sequence 4, Application US/09335409
 ; Patent No. 6121029
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT APPLICATION NUMBER: US/09/335,409
 ; CURRENT FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 1832
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-335-409-4

Query Match 6.3%; Score 79; DB 3; Length 1832;
 Best Local Similarity 26.3%; Pred. No. 5.1;
 Matches 54; Conservative 25; Mismatches 80; Indels 46; Gaps 9;
 QY 19 GAARSVAGGLAE-----SDDALVEQLTSQALPVRACRLVSRSEQETSOLVAR 67

DB 1451 GLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAVAALAEARGARVTYAKAD-----VAD 1504
 QY 68 LEQLFPVSALVASSGKQLAGVRN--GYLD-----YORLGDRWLALVAHHAK 114
 DB 1505 RAQLERILREVTTSGMPLRGVVAAGILDDGLLMQOOPARFKYMAPKVOGALHHLALTR 1564
 QY 115 KACLVIDLGTAVTSDLVAAADGVHGGYICPGMTLMRSQRLTHTRIRYDDAEARRALA-- 172
 DB 1565 EAPLSFFVLVYASGVGLGSPG--QGNYAAANTFL--DALAHRR-----AQGLPALSD 1614
 QY 173 ---SLOPGQATAEAVRGCLLMRLRG 194
 DB 1615 WGLFAEVGMAAAQE-DRGARLVSRG 1638

RESULT 4
 US-09-568-102-4
 ; Sequence 4, Application US/09568102
 ; Patent No. 6346404
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT APPLICATION NUMBER: US/09/568,102
 ; CURRENT FILING DATE: 2000-05-10
 ; PRIOR APPLICATION NUMBER: 09/335,409
 ; PRIOR FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 1832
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-568-102-4

Query Match 6.3%; Score 79; DB 4; Length 1832;
 Best Local Similarity 26.3%; Pred. No. 5.1;
 Matches 54; Conservative 25; Mismatches 80; Indels 46; Gaps 9;
 QY 19 GAARSVAGGLAE-----SDDALVEQLTSQALPVRACRLVSRSEQETSOLVAR 67
 DB 1451 GLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAVAALAEARGARVTYAKAD-----VAD 1504
 QY 68 LEQLFPVSALVASSGKQLAGVRN--GYLD-----YORLGDRWLALVAHHAK 114
 DB 1505 RAQLERILREVTTSGMPLRGVVAAGILDDGLLMQOOPARFKYMAPKVOGALHHLALTR 1564
 QY 115 KACLVIDLGTAVTSDLVAAADGVHGGYICPGMTLMRSQRLTHTRIRYDDAEARRALA-- 172
 DB 1565 EAPLSFFVLVYASGVGLGSPG--QGNYAAANTFL--DALAHRR-----AQGLPALSD 1614
 QY 173 ---SLOPGQATAEAVRGCLLMRLRG 194
 DB 1615 WGLFAEVGMAAAQE-DRGARLVSRG 1638

RESULT 5
 US-09-567-969-4
 ; Sequence 4, Application US/09567969
 ; Patent No. 6355457
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/567,969
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 4
;; LENGTH: 1832
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
US-09-567-969-4

Query Match 6.3%; Score 79; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 5.1;
Matches 54; Conservative 25; Mismatches 80; Indels 46; Gaps 9;
QY 19 GAARSVAGGLAE-----SDDALVEQLTSQALPVRACRLVSVRSQETSQVLAR 67
Db 1451 GLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVALAARGARVTYAKAD-----VAD 1504
QY 68 LEQLFPVSALVASSGKQLAGVRN--GYLD-----YORGLDRWLALVAHAHLAK 114
Db 1505 RAQLERILREVTSGMPLRGVVHAAGILDDGLLMQOTPARFKVMAPKVQAGALHLALTR 1564
QY 115 KACLVIDLGTAVTSDLVAAADGVHLGGYICPGMTLMRSQRLTHTRIRYDDAEARRALA-- 172
Db 1565 EAPLSFFVLYASGVGLGSPG--QGNVAAANTFL---DALAHRR-----AOGLPALSVD 1614
QY 173 ---SLOPGQATAEAVRGCLLMRLG 194
Db 1615 WGLFAEVGMAAAQE--DRGARLVSRG 1638

RESULT 6
US-09-568-480-4
;; Sequence 4, Application US/09568480
;; Patent No. 6355458
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyt, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,480
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 4
;; LENGTH: 1832
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
US-09-568-480-4

Query Match 6.3%; Score 79; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 5.1;
Matches 54; Conservative 25; Mismatches 80; Indels 46; Gaps 9;
QY 19 GAARSVAGGLAE-----SDDALVEQLTSQALPVRACRLVSVRSQETSQVLAR 67
Db 1451 GLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVALAARGARVTYAKAD-----VAD 1504
QY 68 LEQLFPVSALVASSGKQLAGVRN--GYLD-----YORGLDRWLALVAHAHLAK 114
Db 1505 RAQLERILREVTSGMPLRGVVHAAGILDDGLLMQOTPARFKVMAPKVQAGALHLALTR 1564
QY 115 KACLVIDLGTAVTSDLVAAADGVHLGGYICPGMTLMRSQRLTHTRIRYDDAEARRALA-- 172

Db 1565 EAPLSFFVLYASGVGLGSPG--QGNVAAANTFL---DALAHRR-----AOGLPALSVD 1614
QY 173 ---SLOPGQATAEAVRGCLLMRLG 194
Db 1615 WGLFAEVGMAAAQE--DRGARLVSRG 1638
RESULT 7
US-09-568-486-4
;; Sequence 4, Application US/09568486
;; Patent No. 6355459
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyt, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,486
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 4
;; LENGTH: 1832
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
US-09-568-486-4

Query Match 6.3%; Score 79; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 5.1;
Matches 54; Conservative 25; Mismatches 80; Indels 46; Gaps 9;
QY 19 GAARSVAGGLAE-----SDDALVEQLTSQALPVRACRLVSVRSQETSQVLAR 67
Db 1451 GLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVALAARGARVTYAKAD-----VAD 1504
QY 68 LEQLFPVSALVASSGKQLAGVRN--GYLD-----YORGLDRWLALVAHAHLAK 114
Db 1505 RAQLERILREVTSGMPLRGVVHAAGILDDGLLMQOTPARFKVMAPKVQAGALHLALTR 1564
QY 115 KACLVIDLGTAVTSDLVAAADGVHLGGYICPGMTLMRSQRLTHTRIRYDDAEARRALA-- 172
Db 1565 EAPLSFFVLYASGVGLGSPG--QGNVAAANTFL---DALAHRR-----AOGLPALSVD 1614
QY 173 ---SLOPGQATAEAVRGCLLMRLG 194
Db 1615 WGLFAEVGMAAAQE--DRGARLVSRG 1638

RESULT 8
US-09-568-472-4
;; Sequence 4, Application US/09568472
;; Patent No. 6358719
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyt, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,472
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 4
LENGTH: 1832
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-568-472-4

Query Match 6.3%; Score 79; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 5.1;
Matches 54; Conservative 25; Mismatches 80; Indels 46; Gaps 9;

QY 19 GAARSVAGGLAE-----SDDALVEQLTSQOALPVACRLVSVRSEQETSQVLR 67
DB 1451 GLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVALAARGARVTVAKAD-----VAD 1504
QY 68 LEQLFPVSALVASSGKQLAGVRN--GYLD-----YQRLGLDRWLALVAHHAK 114
DB 1505 RAQLERILREVTTSGMPLRGVHAAGILDDGLLMOQTPARFRKVMAPKVOGALHLHALTR 1564
QY 115 KACLVLDGTAVTSDIVAAGVHLGGVYICPGMTLMRSQRLTHRTIRYDDAEARRALA-- 172
DB 1565 EAPLSFVLVYASGVLLGSPG--QGNVAAANTFL---DALAHRR-----AQGLPALSD 1614
QY 173 ---SLOPGQATAEAVRGCLLMRLG 194
DB 1615 WGLFAEVGMAAQE-DRGARLVSRG 1638

RESULT 9
US-09-567-899-4
Sequence 4, Application US/09567899
Patent No. 6383787
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,899
PRIOR FILING DATE: 2000-05-10
PRIORITY FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 4
LENGTH: 1832
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-567-899-4

Query Match 6.3%; Score 79; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 5.1;
Matches 54; Conservative 25; Mismatches 80; Indels 46; Gaps 9;

QY 19 GAARSVAGGLAE-----SDDALVEQLTSQOALPVACRLVSVRSEQETSQVLR 67
DB 1451 GLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVALAARGARVTVAKAD-----VAD 1504
QY 68 LEQLFPVSALVASSGKQLAGVRN--GYLD-----YQRLGLDRWLALVAHHAK 114
DB 1505 RAQLERILREVTTSGMPLRGVHAAGILDDGLLMOQTPARFRKVMAPKVOGALHLHALTR 1564
QY 115 KACLVLDGTAVTSDIVAAGVHLGGVYICPGMTLMRSQRLTHRTIRYDDAEARRALA-- 172
DB 1565 EAPLSFVLVYASGVLLGSPG--QGNVAAANTFL---DALAHRR-----AQGLPALSD 1614
QY 173 ---SLOPGQATAEAVRGCLLMRLG 194
DB 1615 WGLFAEVGMAAQE-DRGARLVSRG 1638

RESULT 10
US-09-413-814-70
Sequence 70, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
TITLE OF INVENTION: heteropolyketide compounds
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 70
LENGTH: 2259
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-70

Query Match 6.3%; Score 79; DB 4; Length 2259;
Best Local Similarity 20.5%; Pred. No. 7.1;
Matches 68; Conservative 27; Mismatches 93; Indels 144; Gaps 13;

QY 16 VIEGAARSVAGGLAESDDALVEQLTSQOALPVACRL-----VSVRSEQETS 62
DB 941 LLDACVQMIYGFADRDGA-----TPWAPVEGVSRLVQSPGELMCHARVVDGQOASS 995
QY 63 QLVARLEQFPVSALVASSGKQL-----AGVRNGYLDYQRLGLD-----RW 103
DB 996 RWSADTFELMDGTGANVAEISRLVVERLASGVRDRDDWFLDWEPAALGGPKITAGR 1055
QY 104 LALVAHHAKKACLV-----DLGTAVTSDLV--AADG-----VHL----- 138
DB 1056 LLLGEGGLGRSLCSALKAAAGHVVAAGDDTSTAGMRALLANAFDQAPTAVVHLSLD 1115
QY 139 -GGYICPGM-----TLMS-----QIRHT----- 157
DB 1116 GGGQGLGGLGAQOALDAPRSPVDADALEALMRGDSVLSVQALVGMDLRNAPRLWL 1175
QY 158 -----RRIRYDDAEARRALASLQPGQATAEAVRGCLLMRLG 194
DB 1176 TRGAQAAAAGDSVQVQAPLGLGRTTALHAELRCISVDLDPAEPEGA----- 1224
QY 195 FVREQVAMACELGPGC--EIFLTGGDAELVR 224
DB 1225 -----DALLAELLADDAEEVALRGDRVAR 1251

RESULT 11
US-09-335-409-7
Sequence 7, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A

;; CURRENT APPLICATION NUMBER: US/09/335,409
;; CURRENT FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 7
;; LENGTH: 2439
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
US-09-335-409-7

Query Match 6.3%; Score 79; DB 3; Length 2439;
Best Local Similarity 20.5%; Pred. No. 8;
Matches 68; Conservative 27; Mismatches 93; Indels 144; Gaps 13;
QY 16 VIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRL-----VSVRSEQETS 62
DB 1121 LLDACVQMVGFADRDDEA-----TPWAPVEGVSVRLFORSGELWCHARVVSDGQQASS 1175
QY 63 QLVARLEQLFPVSALVASSGKQL-----AGVRNGYLDYQRLGLD-----RW 103
DB 1176 RWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLELDWEPAALGGPKITAGRW 1235
QY 104 LALVAHHLAKKACLV-----DLGTAVTSDLV--AADG-----VHL----- 138
DB 1236 LLLGEGGLGRSLCSALKAAAGHVVVHAAGDDTSTAGMRALLANAFDGOAPTAVVHLSLD 1295
QY 139 -GGYICPGM-----TLMRS-----QLRTHT----- 157
DB 1296 GGGQLGPGGALDAPRSPDADALESMRCDSDVLSVQALVGMNDLBNAPRLWLL 1355
QY 158 -----RRIRYDDAEARRALASLOPQOATAEAVRGCLLMRLG 194
DB 1356 TRGAQAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDPAEPEGEA----- 1404
QY 195 FVREYQYMACELLLGPDCE--EIFLTGGDAELVR 224
DB 1405 -----DALLAELLADDAEEVALRGDRILVAR 1431

RESULT 12

US-09-568-102-7
;; Sequence 7, Application US/09568102
;; Patent No. 6346404
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,102
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 7
;; LENGTH: 2439
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
US-09-568-102-7

Query Match 6.3%; Score 79; DB 4; Length 2439;
Best Local Similarity 20.5%; Pred. No. 8;
Matches 68; Conservative 27; Mismatches 93; Indels 144; Gaps 13;
QY 16 VIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRL-----VSVRSEQETS 62
DB 1121 LLDACVQMVGFADRDDEA-----TPWAPVEGVSVRLFORSGELWCHARVVSDGQQASS 1175
QY 63 QLVARLEQLFPVSALVASSGKQL-----AGVRNGYLDYQRLGLD-----RW 103

DB 1176 RWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLELDWEPAALGGPKITAGRW 1235
QY 104 LALVAHHLAKKACLV-----DLGTAVTSDLV--AADG-----VHL----- 138
DB 1236 LLLGEGGLGRSLCSALKAAAGHVVVHAAGDDTSTAGMRALLANAFDGOAPTAVVHLSLD 1295
QY 139 -GGYICPGM-----TLMRS-----QLRTHT----- 157
DB 1296 GGGQLGPGGALDAPRSPDADALESMRCDSDVLSVQALVGMNDLBNAPRLWLL 1355
QY 158 -----RRIRYDDAEARRALASLOPQOATAEAVRGCLLMRLG 194
DB 1356 TRGAQAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDPAEPEGEA----- 1404
QY 195 FVREYQYMACELLLGPDCE--EIFLTGGDAELVR 224
DB 1405 -----DALLAELLADDAEEVALRGDRILVAR 1431

RESULT 13

US-09-567-969-7
;; Sequence 7, Application US/09567969
;; Patent No. 6355457
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/567,969
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 7
;; LENGTH: 2439
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
US-09-567-969-7

Query Match 6.3%; Score 79; DB 4; Length 2439;
Best Local Similarity 20.5%; Pred. No. 8;
Matches 68; Conservative 27; Mismatches 93; Indels 144; Gaps 13;
QY 16 VIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRL-----VSVRSEQETS 62
DB 1121 LLDACVQMVGFADRDDEA-----TPWAPVEGVSVRLFORSGELWCHARVVSDGQQASS 1175
QY 63 QLVARLEQLFPVSALVASSGKQL-----AGVRNGYLDYQRLGLD-----RW 103
DB 1176 RWSADFELMDGTGAVVAEISRLVVERLASGVRRRDADDWFLELDWEPAALGGPKITAGRW 1235
QY 104 LALVAHHLAKKACLV-----DLGTAVTSDLV--AADG-----VHL----- 138
DB 1236 LLLGEGGLGRSLCSALKAAAGHVVVHAAGDDTSTAGMRALLANAFDGOAPTAVVHLSLD 1295
QY 139 -GGYICPGM-----TLMRS-----QLRTHT----- 157
DB 1296 GGGQLGPGGALDAPRSPDADALESMRCDSDVLSVQALVGMNDLBNAPRLWLL 1355
QY 158 -----RRIRYDDAEARRALASLOPQOATAEAVRGCLLMRLG 194
DB 1356 TRGAQAAAGDVSVVQAPLLGLGRTIALEHAELRCISVDLDPAEPEGEA----- 1404
QY 195 FVREYQYMACELLLGPDCE--EIFLTGGDAELVR 224
DB 1405 -----DALLAELLADDAEEVALRGDRILVAR 1431

RESULT 14

US-09-568-480-7
; Sequence 7, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-7

Query Match 6.3%; Score 79; DB 4; Length 2439;
Best Local Similarity 20.5%; Pred. No. 8;
Matches 68; Conservative 27; Mismatches 93; Indels 144; Gaps 13;

Qy 16 VIEGAARSVAGGLAESDDALVQLTSQOALPVRACRL-----VSVRSEQETS 62
Db 1121 LLDACVQMVIGAFADREA-----TPWAPVEGVSRLVFORSPGELWCHWVSDGQASS 1175
Qy 63 OLVARLEQLFPVSALVASSGKQL-----AGVRNGYLDYQRLGLD-----RW 103
Db 1176 RWSADFEIMDGTGAVVAEISRLVVERLASGVRRDADDNFWLELDWEPAALGGPKITAGR 1235
Qy 104 LALVAHHHAKKACLV-----DLGTAVTSDLV--AADG-----VHL----- 138
Db 1236 LLLGEGGIGRSLCSALKAAAGHVVAAGDDTSTAGMRALLANAFDGOAPTAVVHLSSLD 1295
Qy 139 -GGYICPGM-----TLMS-----QIRHT----- 157
Db 1296 GGGQGLGGAQAGDAPSPVDADALESAALMRGCDVSLVQALVGMDLRNAPRLWL 1355
Qy 158 TRGQAAAAGDVSVVQAPLGLGRTIALEHAELRCISVDLDPAEPEGEA----- 1404
Db 1356 FVREQYAMACELIGPDC--EIFLTGGDAELVR 224
1405 -----DALLAELLADDAEEVALRGGRDLVAR 1431

RESULT 15

US-09-568-486-7
; Sequence 7, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-7

Query Match 6.3%; Score 79; DB 4; Length 2439;
Best Local Similarity 20.5%; Pred. No. 8;
Matches 68; Conservative 27; Mismatches 93; Indels 144; Gaps 13;

Qy 16 VIEGAARSVAGGLAESDDALVQLTSQOALPVRACRL-----VSVRSEQETS 62
Db 1121 LLDACVQMVIGAFADREA-----TPWAPVEGVSRLVFORSPGELWCHWVSDGQASS 1175
Qy 63 OLVARLEQLFPVSALVASSGKQL-----AGVRNGYLDYQRLGLD-----RW 103
Db 1176 RWSADFEIMDGTGAVVAEISRLVVERLASGVRRDADDNFWLELDWEPAALGGPKITAGR 1235
Qy 104 LALVAHHHAKKACLV-----DLGTAVTSDLV--AADG-----VHL----- 138
Db 1236 LLLGEGGIGRSLCSALKAAAGHVVAAGDDTSTAGMRALLANAFDGOAPTAVVHLSSLD 1295
Qy 139 -GGYICPGM-----TLMS-----QIRHT----- 157
Db 1296 GGGQGLGGAQAGDAPSPVDADALESAALMRGCDVSLVQALVGMDLRNAPRLWL 1355
Qy 158 TRGQAAAAGDVSVVQAPLGLGRTIALEHAELRCISVDLDPAEPEGEA----- 1404
Db 1356 FVREQYAMACELIGPDC--EIFLTGGDAELVR 224
1405 -----DALLAELLADDAEEVALRGGRDLVAR 1431

Search completed: June 24, 2003, 21:49:32
Job time : 8.24485 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.307 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-20
Perfect score: 1248
Sequence: 1 MILEDCGNSLIKRWVIEGA.....GARIMPOLVFVGLALACPIE 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 41779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	1248	100.0	248	9	US-09-813-453A-20
2	794.5	63.7	249	9	US-09-813-453A-61
3	789.5	63.3	249	9	US-09-813-453A-70
4	234.5	18.8	592	9	US-09-813-453A-43
5	232.5	18.6	460	9	US-09-813-453A-39
6	231.5	18.5	592	9	US-09-813-453A-22
7	195.5	15.7	244	9	US-09-813-453A-41
8	193	15.5	242	9	US-09-813-453A-65
9	178.5	14.3	267	9	US-09-813-453A-15
10	160	12.8	260	9	US-09-813-453A-51
11	159.5	12.8	223	9	US-09-895-913A-74
12	159.5	12.8	223	9	US-09-813-453A-14
13	159.5	12.8	223	9	US-09-813-453A-67
14	159.5	12.8	250	9	US-09-813-453A-3
15	157	12.6	257	9	US-09-813-453A-53
16	149	11.9	272	9	US-09-813-453A-5
17	147	11.8	272	9	US-09-712-363-276
18	145	11.6	246	9	US-09-813-453A-9
19	144.5	11.6	241	9	US-09-813-453A-63

20	141	11.3	255	9	US-09-813-453A-7	Sequence 7, Appli
21	138.5	11.1	265	9	US-09-813-453A-4	Sequence 4, Appli
22	136	10.9	219	9	US-09-813-453A-57	Sequence 57, Appli
23	136	10.9	258	9	US-09-813-453A-6	Sequence 6, Appli
24	136	10.9	262	9	US-09-813-453A-45	Sequence 45, Appli
25	136	10.9	273	9	US-09-813-453A-10	Sequence 10, Appli
26	133	10.7	258	9	US-09-813-453A-2	Sequence 2, Appli
27	123.5	9.9	258	9	US-09-813-453A-49	Sequence 49, Appli
28	121	9.7	254	9	US-09-813-453A-47	Sequence 47, Appli
29	120.5	9.7	209	9	US-09-813-453A-21	Sequence 21, Appli
30	119.5	9.6	256	9	US-09-813-453A-55	Sequence 55, Appli
31	111	8.9	233	9	US-09-813-453A-17	Sequence 17, Appli
32	107	8.6	212	9	US-09-813-453A-59	Sequence 59, Appli
33	107	8.6	257	9	US-09-813-453A-13	Sequence 13, Appli
34	98.5	7.9	229	9	US-09-813-453A-12	Sequence 12, Appli
35	89	7.1	4150	9	US-09-808-880-2	Sequence 2, Appli
36	88.5	7.1	277	9	US-09-738-626-4512	Sequence 4512, Ap
37	86	6.9	277	9	US-09-738-626-4732	Sequence 4732, Ap
38	85	6.8	298	9	US-09-712-363-253	Sequence 253, App
39	84	6.7	956	10	US-09-815-242-11925	Sequence 11925, A
40	84	6.7	1827	9	US-09-712-363-261	Sequence 261, App
41	83.5	6.7	262	9	US-09-813-453A-11	Sequence 11, Appli
42	83	6.7	392	10	US-09-945-825-8	Sequence 8, Appli
43	83	6.7	467	10	US-09-841-880-4	Sequence 4, Appli
44	82.5	6.6	251	10	US-09-815-242-12053	Sequence 12053, A
45	82.5	6.6	258	9	US-09-712-363-220	Sequence 220, App

ALIGNMENTS

RESULT 1
US-09-813-453A-20
; Sequence 20, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-813-453A-20

Query Match	100.0%	Score 1248;	DB 9;	Length 248;
Best Local Similarity	100.0%	Pred. No. 1.3e-117;		
Matches 248;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MILEDCGNSLIKRWVIEGAARSVAGGLAESD	1	MILEDCGNSLIKRWVIEGAARSVAGGLAESD
DB	1	MILEDCGNSLIKRWVIEGAARSVAGGLAESD	1	MILEDCGNSLIKRWVIEGAARSVAGGLAESD
QY	61	TSQVLRLEQLFPVSALVASSGKQLAGVNGVLDYQRLDRWLVAHHLAKKACLV	120	TSQVLRLEQLFPVSALVASSGKQLAGVNGVLDYQRLDRWLVAHHLAKKACLV
DB	61	TSQVLRLEQLFPVSALVASSGKQLAGVNGVLDYQRLDRWLVAHHLAKKACLV	120	TSQVLRLEQLFPVSALVASSGKQLAGVNGVLDYQRLDRWLVAHHLAKKACLV
QY	121	DLGTAVTSLVAADGVHLGGYICPGMTLMRSOLTRTHRRYDDAEARRALASLOPQAT	180	DLGTAVTSLVAADGVHLGGYICPGMTLMRSOLTRTHRRYDDAEARRALASLOPQAT
DB	121	DLGTAVTSLVAADGVHLGGYICPGMTLMRSOLTRTHRRYDDAEARRALASLOPQAT	180	DLGTAVTSLVAADGVHLGGYICPGMTLMRSOLTRTHRRYDDAEARRALASLOPQAT
QY	181	AEAVRGCLMLRGFVREQYAMACELLGPDCEFTTGGDAELVRDELAGARIMPLVFG	240	AEAVRGCLMLRGFVREQYAMACELLGPDCEFTTGGDAELVRDELAGARIMPLVFG
DB	181	AEAVRGCLMLRGFVREQYAMACELLGPDCEFTTGGDAELVRDELAGARIMPLVFG	240	AEAVRGCLMLRGFVREQYAMACELLGPDCEFTTGGDAELVRDELAGARIMPLVFG

Db 507 TTGNVAVSGMMDAVCGSVMMHGRLEKKTG-----AGKPDVDTVTITGGGAAKVAEALPPAF 561
QY 231 -----RIMPDVLFVGL 241
Db 562 LAENTVRVADNLVIYGL 578

RESULT 5

US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-813-453A-39

Query Match 18.6%; Score 232.5; DB 9; Length 460;
Best Local Similarity 27.2%; Pred. No. 4.5e-15;
Matches 70; Conservative 41; Mismatches 109; Indels 37; Gaps 7;
QY 3 LEIDCGNSLIKRWVIEGARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
Db 209 LLEGGNSRLKAWWENGTFATVGSAPYRD---LSPGLAEAEKVDGNVRIVGCAVCGESKA 268
QY 60 ETSOLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGDRWLALVAHHAKKACLV 119
Db 269 QVKEQLARKIEWLPSSA-----QALGIRNHYRHPHEGSDRWFNALGSRFRNACV 321
QY 120 IDLGTAVTSDLVAAADVHGGYICPGMTLMRSQRLTHTTRIRYDDAEARRALASLP--- 176
Db 322 VSCGTATVDALDDGHYLGTTMPGFLMKESLAVT-----ANLNRPAKRYPFPT 374
QY 177 --GOATAEAVER---GCLMLRGFVREQYAMACELGPDCEIFLTGGDAELVRDELAGA- 230
Db 375 TTGNVAVSGMMDAVCGSVMMHGRLEKKTG-----AGKPDVDTVTITGGGAAKVAEALPPAF 429
QY 231 -----RIMPDVLFVGL 241
Db 430 LAENTVRVADNLVIHGL 446

RESULT 6

US-09-813-453A-22
; Sequence 22, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-813-453A-22

Query Match 18.5%; Score 231.5; DB 9; Length 592;
Best Local Similarity 27.7%; Pred. No. 8.1e-15;
Matches 72; Conservative 38; Mismatches 107; Indels 43; Gaps 8;
QY 3 LEIDCGNSLIKRWVIEGARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVR 56
Db 341 LLLDGGNSRLKAWWENGTFATVGSAPYRD---LSPGLAEAEKVDGNVRIVGCAVCGEF 397
QY 57 SEQETSOLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGDRWLALVAHHAKK 116
Db 398 KKAQVOEQLARKIEWLPSSA-----QALGIRNHYRHPHEGSDRWFNALGSRFRSNA 450
QY 117 CLVIDLGTAVTSDLVAAADVHGGYICPGMTLMRSQRLTHTTRIRYDDAEARRALASLP 176
Db 451 CVVSCGTATVDALDDGHYLGTTMPGFLMKESLAVT-----ANLNRHAGKRYP 503
QY 177 ---GOATAEAVER---GCLMLRGFVREQYAMACELGPDCEIFLTGGDAELVRDELA 228
Db 504 FPTTGNVAVSGMMDAVCGSVMMHGRLEKKTG-----AGKPDVDTVTITGGGAAKVAEALP 558
QY 229 GA-----RIMPDVLFVGL 241
Db 559 PAFLAENTVRVADNLVIHGL 578

RESULT 7

US-09-813-453A-41
; Sequence 41, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-813-453A-41

Query Match 15.7%; Score 195.5; DB 9; Length 244;
Best Local Similarity 30.4%; Pred. No. 1e-11;
Matches 80; Conservative 33; Mismatches 95; Indels 55; Gaps 12;
QY 3 LEIDCGNSLIKRWVIEGARSVA---GGLAESDDALV-----EQLTSQALP----- 46
Db 5 LIVDQGNACKAVFVRNNSIESIFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDERAEA 64
QY 47 -VRACRLVSVRSEQETSOLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGDRWL 105
Db 65 IVRSCAAASLMMGTTET-----PVP-----LRLQY-DRRTLGADRLAA 100
QY 106 LVAHHIL-AKKACLVLDLGTAVTSDLVAAADVHGGYICPGMTLMRSQRLTHTTRIRYDD 164
Db 101 VGAHSLYPNTELLVIDAGTAITYRVSAGIYGGNISPLGHLRKFALHLFTGRUPLID 160

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 250

; TYPE: PRF

; ORGANISM: Clostridium acetobutylicum

US-09-813-453A-3

Query Match 12.8%; Score 159.5; DB 9; Length 250;

Best Local Similarity 26.1%; Pred. No. 4.3e-08;

Matches 65; Conservative 40; Mismatches 101; Indels 43; Gaps 9;

Qy 1 MILEDCGNSLI-----KWRVIEGAARSVAGGLAESDDALVEQLTSQOALPVR 48

Db 18 VILVDVGNINVLGIYNDTKLTAEWRLSTDVLS-----ADEYGIQVMNLFQODKLDPT 72

Qy 49 ACRUVSVRSQETSQVLRLEQL-----FVPSALVASSGKOLAGVRNGYLDYQRLGLDRWL 104

Db 73 LVEGVIIIS--VVPNIMYSLEHMIRKYKINPLVVGPIR-TGINIKYDNPKEVGADRIV 129

Qy 105 ALVAAHHLAKACLVLDGTAVTSDLVADGVHLGGYICPGMTLMRSQLRTHTRIRYDD 164

Db 130 NVAHAEIYKSLIIDFGTATTCVAVRENGDYLGGACIGKVSSEAL-----FEK 181

Qy 165 AEARRALASLPQGA-----TAEAVERGCLLMRLRGFVREQYAMACELIGPDCE-----IF 214

Db 182 AAKLPVELIKPAPYAIKNTISSIQSIGIVRYRLQVKYLFKELKENL-PDGRRTTSLVL 240

Qy 215 LTGGDAELV 223

Db 241 ATGGLAKLI 249

RESULT 15

US-09-813-453A-53

; Sequence 53; Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 53

; LENGTH: 257

; TYPE: PRF

; ORGANISM: Chlorobium tepidum

US-09-813-453A-53

Query Match 12.6%; Score 157; DB 9; Length 257;

Best Local Similarity 24.7%; Pred. No. 8e-08;

Matches 63; Conservative 49; Mismatches 111; Indels 32; Gaps 10;

Qy 1 MILEDCGNSLIKWRVIEG-----AARSVAGGLAESDDALVEQLTSQOALPVRACRLVSVR 56

Db 1 MRLVVDIGNTTTTLAIFTGDEEPSVESVPSALF-ADSSMTREVFNGNMARKHGEPOATAIC 59

Qy 57 SEQETSOLV--ARLEOLFVPSALVASSGKOLAGVRNGYLDYQRLGLDRWLALVA--AHHL 112

Db 60 SVVPSATAVGSALLESLSVPLTICC-KLRFPPRLDYATPHTFGADR-LALCAWSRHLF 117

Qy 113 AKKACLVIDLGTAVTSDLVADGVHLGGYICPGMTLMRSQLRTHTR---RIRYDDAEARR 169

Db 118 SERPVIADVIGTAITFDVLDVTGNYRGGLIMPGLMMAGALHSRTAQLPQVIRIDRPESLL 177

Qy 170 ALASLOPQOATAEAVERGCLLMRLRGFVREQYAMACELLG-----PDCEIFLTGDAE 221

Db 178 -----GRSTTECIKSG---VFVGVVVKQIGGLVDAIRGDLVRDFGCESTVEIVITGNSR 227

Qy 222 LVRDELAGARIMPD 236

Db 228 IIVPEIGPVSVIDE 242

Search completed: June 24, 2003, 22:24:04

Job time : 12.307 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.3857 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-13
Perfect score: 1368
Sequence: 1 METSKPGCGALNDKQKPW.....DDNLIFLGMAAIHHGDRPIC 257

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1368	100.0	257	23 AAU91160	Synechocystis pant
2	174	12.7	241	23 AAU91179	Pantothenate kinas
3	174	12.7	265	23 AAU91151	Streptomyces coeli
4	170	12.4	259	23 ABB47661	Listeria monocytog
5	163	11.9	258	22 AAU01243	B. subtilis novel
6	163	11.9	258	23 AAU91149	Bacillus subtilis
7	158	11.5	233	23 AAU91163	Pantothenate kinas
8	156.5	11.4	262	23 AAU91170	Pantothenate kinas
9	152	11.1	244	23 AAU91168	Pantothenate kinas
10	150	11.0	258	23 AAU91172	Pantothenate kinas

us-09-813-453a-13.rag

Page 1

11 147 10.7 260 23 AAU91173

12 146.5 10.7 257 23 AAU91174

13 144.5 10.6 256 23 AAU91175

14 144 10.5 219 23 AAU91176

15 141.5 10.3 223 19 AAU98422

16 141.5 10.3 223 23 AAU91161

17 141.5 10.3 223 23 AAU91181

18 141 10.3 455 21 AAU74910

19 141 10.3 592 20 AAU38615

20 141 10.3 592 20 AAU38616

21 141 10.3 592 21 AAU74912

22 141 10.3 592 21 AAU74913

23 141 10.3 592 23 AAU91166

24 141 10.3 592 23 AAU91169

25 138 10.1 246 23 AAU91156

26 137.5 10.1 254 23 AAU91171

27 137 10.0 455 20 AAU38617

28 137 10.0 455 21 AAU74908

29 137 10.0 460 23 AAU91167

30 137 10.0 592 20 AAU38618

31 137 10.0 592 21 AAU74911

32 136.5 10.0 273 23 AAU91157

33 136.5 10.0 389 21 AAU74909

34 133 9.7 229 23 AAU91159

35 126 9.2 255 23 AAU91154

36 123.5 9.0 262 23 AAU91155

37 121.5 8.9 250 23 AAU91150

38 118 8.6 249 23 AAU91178

39 117 8.6 258 23 AAU91153

40 116.5 8.5 209 23 AAU91165

41 112.5 8.2 249 23 AAU91182

42 107 7.8 248 23 AAU91164

43 107 7.8 521 22 AAU58846

44 106 7.7 272 22 AAG81225

45 106 7.7 272 23 AAU91152

ALIGNMENTS

RESULT 1

AAU91160

ID AAU91160 standard; Protein; 257 AA.

XX

AC AAU91160;

XX

DT 05-JUN-2002 (first entry)

XX

DE Synechocystis pantothenate kinase Coax.

XX

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX

OS Synechocystis sp.

XX

PN WO200216601-A2.

XX

PD 28-FEB-2002.

XX

PF 24-AUG-2001; 2001WO-US26531.

XX

PR 24-AUG-2000; 2000US-227860P.

XX

PR 20-MAR-2001; 2001US-0813453.

XX

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX

PI Yocum RR, Patterson TA;

XX

DR WPI; 2002-269358/31.

XX

PT Identifying potential antibiotic or antimicrobial agent, comprises

XX

PT contacting composition comprising pantothenate kinase (Coax) protein

XX

PT with test compound and identifying inhibitor of the Coax protein -

xx	Claim 10; Page 77-78; 128pp; English.
xx	The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.
xx	
SQ	Sequence 257 AA;
	Query Match 100.0%; Score 1368; DB 23; Length 257;
	Best Local Similarity 100.0%; Pred. No. 6.7e-122;
	Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
	1 METSPGGCLALDNDKQKPWLGLMGINSRLHWAYCSGNAPLOTWTWTDYNPKSQALPVLIG 60 1 METSPGGCLALDNDKQKPWLGLMGINSRLHWAYCSGNAPLOTWTWTDYNPKSQALPVLIG 60
Qy	61 KVPMLASVYPQEETWVRVYOPKILTLKNLPVLNLYPSFGIDRALAGLTGTYGPPCLV 120
Db	61 KVPMLASVYPQEETWVRVYOPKILTLKNLPVLNLYPSFGIDRALAGLTGTYGPPCLV 120 121 VDGSGTALTITGFDQDKLVGAAILPGILGLOLATLGDRLAALPKLEMDQLTELDPDRWALDT 180 121 VDGSGTALTITGFDQDKLVGAAILPGILGLOLATLGDRLAALPKLEMDQLTELDPDRWALDT 180
Qy	181 PSAIFSGVVYGVILGALSQYLQDWOKLFPGAAMVITGGDGKILHGFLKEHSPNLISVAWDNN 240
Db	181 PSAIFSGVVYGVILGALSQYLQDWOKLFPGAAMVITGGDGKILHGFLKEHSPNLISVAWDNN 240 241 LIFLGMAATHHGDRPIC 257
Qy	241 LIFLGMAATHHGDRPIC 257 241 LIFLGMAATHHGDRPIC 257
Db	241 LIFLGMAATHHGDRPIC 257 241 LIFLGMAATHHGDRPIC 257

RESULT 2	
AAU911179	
ID	AAU911179 standard; Protein; 241 AA.
XX	
AC	AAU911179;
XX	
DT	05-JUN-2002 (first entry)
XX	
XX	Pantothenate kinase (CoaX) #17.
KW	Pantothenate kinase; CoaX; antibiotic; antimicrobial;
XX	pantothenate kinase modulator; coenzyme A; bactericidal compound.
OS	Thiobacillus ferrooxidans.
XX	
PN	WO200216601-A2.
PD	28-FEB-2002.
PF	
XX	24-AUG-2001; 2001WO-US26531.
XX	
PR	24-AUG-2000; 2000US-237860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WPI; 2002-269358/31.
DR	N-PSDB; ABR54200.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprises

PT	contacting composition comprising pantothenate kinase (Coax) protein
PT	with test compound and identifying inhibitor of the Coax protein -
XX	
PS	Claim 10; Page 109-110; 128pp; English.
XX	
CC	The invention describes assays for identifying a (potential) antibiotic
CC	comprising contacting an assay composition comprising a pantothenate
CC	kinase (Coax) protein with a test compound, and determining the ability
CC	of the test compound to inhibit the activity of the Coax protein, an
CC	essential enzyme for the production of coenzyme A. Coax protein is a
CC	valuable target for identifying bactericidal compounds. Coax modulating
CC	agents can be used in an infectious animal model to determine the
CC	efficacy, toxicity, or side effects of treatment with such an agent. This
CC	is the amino acid sequence of a pantothenate kinase (Coax) protein
CC	described in the invention.
XX	
SQ	Sequence 241 AA;
	Query Match 12.7%; Score 174; DB 23; Length 241;
	Best Local Similarity 27.6%; Pred.No. 2.5e-08;
	Matches 60; Conservative 34; Mismatches 71; Indels 52; Gaps 8;
QY	40 PLOTWVTDYNPKSAQLPVLGKGVPLMLASVVPQTEVWR-----VYQPKILTILKNLP 92
Db	I : I I I : : : : I I I : : : : I : : : : :
	40 PGLTWLSAPN-----REPVALGGVVPAALAAAREALATAEVREPDPGFFERRAVP 88
QY	93 VNIYP--SFGIDRALAGLTGLTY-GFPCLVVDGGTALTITGFDQDKLV-----GGAI 143
Db	: : : : I I I : : : : I : : : : I I I : : : : I : : : : :
	89 HDYHPESLGDFDRRCCLLAAMDYPGODSIDVMGTAITI-----DLLAGGHFRGGRI 141
QY	144 LPGIGLQAFLGRALALPKLEMDQLTELDPDRWALDTPSAIFSGVVVGVLGAQSILQDW 203
Db	I I I : : : : I I I : : : : I : : : : I I I : : : : I : : : : :
	142 LPGIAMSLRGHEGTALLPEVNLNAPAEMIGN--DTSNAIQAGVIHLFADALRGAITDF 198
QY	204 QKLFPGAAMVITGCDGK-----ILHGF 225
Db	: : : : I I I I I : : : : I : : : : I I I : : : : I : : : : :
	199 RQYSQPARIILITGDAPERPWPQGIAGSLYQPHLLLRGF 235
RESULT 3	
AAU91151	
ID	AAU91151 standard; Protein; 265 AA.
XX	
AC	AAU91151;
XX	
DT	05-JUN-2002 (first entry)
XX	
DE	Streptomyces coelicolor pantothenate kinase Coax.
KW	Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.
OS	Streptomyces coelicolor.
XX	
PN	WO200216601-A2.
XX	
PD	28-FEB-2002.
XX	
PF	24-AUG-2001; 2001WO-US26531.
XX	
PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WFI; 2002-269358/31.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprises
PT	contacting composition comprising pantothenate kinase (coax) protein
PT	with test compound and identifying inhibitor of the Coax protein -
XX	

Claim 10; Page 69-70; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 265 AA;

Query Match 12.7%; Score 174; DB 23; Length 265;

Best Local Similarity 24.9%; Pred. No. 2.9e-08;

Matches 67; Conservative 45; Mismatches 105; Indels 52; Gaps 11;

21 LGLMIGNRLHWAYCSGNAPLQW--VTDYNPKSAQLPVLL-----GRVPLM----- 65

3 LTIDVGNTHVLGLDFGEDIHWRISTDSRRTADELAVLLQGLMGHPHLLGDELGDGID 62

66 -----LASVPEFTEWRY-----QPKLTILKNPLVNLVP-SFGIDRALAGL 108

63 GIAICATVPSVLHREVTTRYDPAVLVEPGVKT--GVPILTDHPKEVGADRIINAV 120

109 GRLGYGPPCLVVDGGTATTGFDQDKLVGGAILPGLQLQATLGDRLAALPKLEMDQ 168

121 RAVELYGGPAIVVDGTTATFDASARGEYIGGVAPGIEISVEALGVKGQALRIEVAR 180

169 LTELPRWALDTPSAIFSGVYVGLQALQSYLDQWQKLF----PGAAMVITGGDGKILHG 224

181 PRSVIGK---NTVEAMQSGIVYGAGQVDGVVNRNARELADDDPDVTVIATGGLAPWVLG 237

225 ---FLKHSPLNSVANDNLIPLGMAATH 250

238 ESSVIDEHEP-----W---LTLMLGLRLVY 258

RESULT 4

ABB47661

ID ABB47661 standard; Protein; 259 AA.

XX

AC ABB47661;

XX

DT 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #365.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX

PD 18-OCT-2001.

XX

PF 11-APR-2001; 2001WO-FR01118.

XX

PR 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

XX

XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehlant J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;

XX

DR WPI; 2002-010914/01.

XX

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides

XX

PS Claim 6; SEQ ID No 366; 192pp; French.

XX

CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies. Identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 259 AA;

Query Match 12.4%; Score 170; DB 23; Length 259;

Best Local Similarity 27.2%; Pred. No. 6.7e-08;

Matches 56; Conservative 43; Mismatches 83; Indels 24; Gaps 8;

QY 64 LMLASVPE-----QTEWVRVQ-----PKILTKNPLVNLVP-SFGIDRALAGLGT 110

DB 60 IITSSVPPIMHAMETMCVFNRIPLIVGPGIKTGLNKLVDN--PREIGSDRIVNAVAA 117

QY 111 GLTYGFPCLVVDGGTALTITGFDQDKLVGGAILPGLQLQATLGDRLAALPKLEMDQLT 170

DB 118 SEETGTPVIVDFGTATTFCYIDESGVYGGGAIAPGIMISTEALYNRAAKLPKRVDAESS 177

QY 171 ELPRWALDTPSAIFSGVYVGLQALQSYLDQWQKLFPGAAMVI-TGGDGKILHGFLEKH 229

DB 178 QIIGK---STVSSMQAGIFYGFGQCEGIIAEMKQSNASPVVATGG----LARMITEK 230

QY 230 SPNLSVANDNLIPLGMAAHHGDRP 255

DB 231 SSAVDIL-DPFLTLKGLLELLYRRNKP 255

RESULT 5

AAU01243

ID AAU01243 standard; Protein; 258 AA.

XX

AC AAU01243;

XX

DT 18-JUL-2001 (first entry)

XX

DE B. subtilis novel pantothenate kinase encoded by the gene coax.
XX Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;
KW nutritional supplement; panto-compound; pantoate.

XX

OS Bacillus subtilis.

XX

PN WO200121772-A2.

XX

PD 29-MAR-2001.

XX

PF 21-SEP-2000; 2000WO-US25993.

XX

PR 21-SEP-1999; 99US-0400494.

20-MAR-2001; 2001US-0813453.
(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.
N-PSDB; ABK54169.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

Disclosure; Page 81-82; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 233 AA;

Query Match 11.5%; Score 158; DB 23; Length 233;

Best Local Similarity 30.6%; Pred. No. 8e-07;

Matches 48; Conservative 34; Mismatches 53; Indels 22; Gaps 7;

64 IMLASVPE-----QTEVRYVQPKIL-----TLKNPLVNLVPSFGIDRALAGLT 110

60 IIIISVVPPINFALERWCTKYFHI-EQIVGPGMKTGLNKKYDN-PREVGADRVNAVAA 117

111 GLTGVPCLVGDTALTITGDDKKLVGGAILPGLGLQLATLGDLAALPKLEMDQLT 170

118 IHLXGNPLVDFGTTATTCYIDENKQYMGAIAPGITISTEALYSRAAKLPRIETIR-- 175

171 ELPDR-WALDTPSAIFSGVYVGLGALQSYLQ--DWQ 204

176 --PDNIGKNTVSAWQSGILFGYGVQEGIVKRMKWQ 210

RESULT 8

AAU91170

AAU91170 standard; Protein; 262 AA.

AAU91170;

05-JUN-2002 (first entry)

Pantothenate kinase (Coax) #8.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Bacillus anthracis.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.
N-PSDB; ABK54191.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

Claim 8; Page 98-99; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 262 AA;

Query Match 11.4%; Score 156.5; DB 23; Length 262;

Best Local Similarity 31.8%; Pred. No. 1.3e-06;

Matches 48; Conservative 32; Mismatches 54; Indels 17; Gaps 7;

79 VYQPKILTLLKPLVNLVLP-SFGIDR---ALAGLTGTLTYGFPCLVDPGGTALTITGDDQ 134

87 VVGFGIKTGLNKKYEN--PREVGADRVNAVAGTH---LYGSPLIIVDFGTTATTCYINE 141

135 DKKLVGGAILPGLGLQLATLGDLAALPKLEMDQLTLPDRWALDTPSAIFSGVYVGLG 194

142 EKHYMGVITPGIMISAEALYSRAAKLPRIETKPPSSVVGK---NTVSAMQSGILYGVVG 198

195 ALQSY---LQDWQKLFPGAAVINGGDKIL 222

199 QVEGIVKRMKEEAKQEP--KVIATGGGLAKLI 227

RESULT 9

AAU91168

AAU91168 standard; Protein; 244 AA.

AAU91168;

05-JUN-2002 (first entry)

Pantothenate kinase (Coax) #6.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Porphyromonas gingivalis.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

N-PSDB; ABK54189.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

XX Claim 8; Page 94-95; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX Sequence 244 AA;

Query Match 11.1%; Score 152; DB 23; Length 244;

Best Local Similarity 27.5%; Pred. No. 3.2e-06;

Matches 42; Conservative 28; Mismatches 77; Indels 6; Gaps 2;

98 SFGIDRALAGLTGLTY-GPPCLVVDGDTALTITGFDQDKKLVGGAILPGLGLQATLGD 156

92 TLGADRLAAVGAHSLYPNTELLVIDAGTAITVERVSAEGIYLGGNISPGHLRFRALHL 151

157 RLAAALPKLEMDQLTELPDRWALDTPSAIFSGVVGVLGALQSLYLDQWQKLFPGCAAMV 216

152 FTGRLPLIDPSGISPKIAEYGSSTEEAITAGVTHGLAGEIDRYIDDLHAKEGRSAVILTG 211

217 GDGKILHGFKEHSPNLSVAWDDNLIFLGMMAI 249

212 GDANYLARIIRS-----GILIHFDVLVLGLNRI 239

RESULT 10

AAU91172

ID AAU91172 standard; Protein; 258 AA.

AC AAU91172;

XX 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #10.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Bacillus stearothermophilus.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX N-PSDB; ABK34193.

XX Identifying potential antibiotic or antimicrobial agent, comprises

XX contacting composition comprising pantothenate kinase (Coax) protein

XX with test compound and identifying inhibitor of the Coax protein

XX Claim 10; Page 101-102; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX Sequence 258 AA;

Query Match 11.0%; Score 150; DB 23; Length 258;

Best Local Similarity 33.1%; Pred. No. 5.3e-06;

Matches 41; Conservative 30; Mismatches 39; Indels 14; Gaps 5;

100 GIDR---ALAGLTGTGLTYGPPCLVVDGDTALTITGFDQDKKLVGGAILPGLGLQATLGD 156

107 GADRIVNAVAGIH---LYGSPLIIVDFGTATTVCYINEHKQYMGGAIPGIMISTEALFA 163

157 RLAAALPKLEMDQLTELPDRWALDTPSAIFSGVVGVLGALQ---SYLDQWQKLFPGCAAMV 213

164 RAAKLPRIEIARDDIIGK---NTVSAMQAGILYGVGVQVEGIVSRMAKSKIPP--KVI 218

214 ITGG 217

219 ATGG 222

RESULT 11

AAU91173

ID AAU91173 standard; Protein; 260 AA.

XX AAU91173;

XX 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #11.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Caulobacter crescentus.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX N-PSDB; ABK54194.

XX Identifying potential antibiotic or antimicrobial agent, comprises

XX contacting composition comprising pantothenate kinase (Coax) protein

XX with test compound and identifying inhibitor of the Coax protein

XX Claim 10; Page 102-103; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

CC described in the invention.

XX Sequence 260 AA;

Query Match 10.7%; Score 147; DB 23; Length 260;

Best Local Similarity 24.1%; Pred. No. 1e-05;

Matches 65; Conservative 41; Mismatches 104; Indels 60; Gaps 11;

21 LGLMIGNSRLHWAYCSGNAPLQTW--VTDYNPKSAQLPVLGKV-----PML 66

3 LAIEGNTNTWFAIHGASWAQWRSATSTRTADEYVWLSQLLSMOGLGFRADAVII 62

67 ASVPEQTEVWRVQPKILTILKNL-----PLV-----NLVPSFGIDRAL 105

63 SSVVPO-----SIFNLNLSRRYFNVEPLVIGENAKLIGDIVRIEKPSEAGADRLV 112

106 AGLGFLTYGPPCLVVDGGTALTITGPDQDKLVGGAILPGLGLQATLGDRLAALPKLE 165

113 NAIGNAWYVPLVVIDSGTATFDIVADCAFEGETIAPGNLSMQALHEAAKLPIRIA 172

166 MDQLTELPDR-WALDTPSAIFSGVYVGLGALQSYLDQWQKLFPGAAVVI--TGGDGKIL 222

173 IQR--PAGNRIVGDTVSAMQSGVFWGVISLIEG-LVARIKAERGEPMVTIATGVSALF 229

223 HGFLK--PHSPNLVAVDDNLIIFLGMAAIH 250

230 EGATDSIDH-----FDSLELTIRGLLEIY 252

RESULT 12

AAU91174 standard; Protein; 257 AA.

AAU91174;

05-JUN-2002 (first entry)

Pantothenate kinase (Coax) #12.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Chlorobium tepidum.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

N-PSDB; ABK54195.

Identifying potential antibiotic or antimicrobial agent, comprises

contacting composition comprising pantothenate kinase (Coax) protein

with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 103-104; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

XX Sequence 257 AA;

Query Match 10.7%; Score 146.5; DB 23; Length 257;

Best Local Similarity 28.3%; Pred. No. 1.1e-05;

Matches 51; Conservative 32; Mismatches 70; Indels 27; Gaps 7;

64 LMLASVPEQTEVWRVQPKILTILKNLPLVNL-----YP-----SFGIDR-ALAGL 108

56 IATCSWVPSATAVGSAL---LESLSFVPVLTICCKLRFFRDYATPHTFGADRLALCAW 112

109 GTGTYGPPCLVVDGGTALTITGPDQDKLVGGAILPGLGLQATLGDRLAALPKLEMDQ 168

113 SRHLESEKPVTAVIDGTAITFDVLTGVNRYGGLIMPGLDMAGALHSRTAQLPQVRDR 172

169 LTELDPDRWALDTPSAIFSGVYVGL-----GALQSYLDQWQKLFPGAA--MVTGGDGKIL 222

173 PESLLGR---STTECIKSGVFWGVYQIGGLVDAIRGDLVRDGFGEVTVITGNSRII 229

RESULT 13

AAU91175 standard; Protein; 256 AA.

AAU91175;

05-JUN-2002 (first entry)

Pantothenate kinase (Coax) #13.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Clostridium difficile.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

N-PSDB; ABK54196.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

Claim 6; Page 105; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

XX Sequence 256 AA;

Query Match 10.6%; Score 144.5; DB 23; Length 256;
Best Local Similarity 23.6%; Pred. No. 1.8e-05;
Matches 61; Conservative 44; Mismatches 108; Indels 45; Gaps 10;
QY 25 IGNSRLHWAYCSGNAPLOTW--VTDYNPKSAQLPVLGKV-----PLMLASVV 70
Db 7 VGNTHMVLGIYKGDVLYNRIKTDREKTSDEYGLISNLFNDVNMISDIDVLISSV 66
QY 71 PEQTEVWR-----VYQPKILTLLKNLPLNLYPSFGIDR---ALAGLGTGLTYG 115
Db 67 PNYHSLNENCIKYCKKQPLIVGPKTKTGLNIIKYN-PKOVGADRVNAVAGIE---KYG 122
QY 116 FPLVDGGRALITGFDQDKLVGGAILPGLGLQLATLGDRLAALPKLEMDQLTLPDR 175
Db 123 APSILVDGRTATTCFAISERKEVYLGTTIAPGKISSSEALFQSASKLPRVE---LAKPGMT 179
QY 176 WALDTPSAIFSGVYVGLGALQSLQDWQKLF--PGNAMVITGGDGKILHGFLEHSPNL 233
Db 180 ICKSTVSAMQSGIYGVGLVDKIIISIMKELNCDDVKVIATGGLAKLI-----ASET 233
234 SVANDONLIEL-GMAAIIH 250
234 SIDYVDGFLTEGLRIIY 251

RESULT 14
AAU91176
ID AAU91176 standard; Protein; 219 AA.
XX AC AAU91176;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #14.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Dehalococcoides ethenogenes.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX DR N-PSDB; ABK34197.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -
XX PS Claim 10; Page 106-107; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.

SQ Sequence 219 AA;
Query Match 10.5%; Score 144; DB 23; Length 219;
Best Local Similarity 23.4%; Pred. No. 1.6e-05;
Matches 50; Conservative 37; Mismatches 75; Indels 52; Gaps 7;
QY 25 IGNSRLHWAYCSGNAPLOTW-----VTDYNPKSAQLPVLGKVP-----L 64
Db 10 IGNTSVNIGIFFEKELLANHLSVAQMADEY-----ASLLGLLQHAGIHPEELNRV 63
QY 65 MLASVPEQT-----EWRVYQPKILTKNLPIV-----NLYPSFGIDRALAG 107
Db 64 IMCSVVPPLTTTPEEFKSY-----FKAPLVVAGIKSVKVRMDNPREVGADRVNA 117
QY 108 LGTGLTYGPFCLVVDGSGTALTITGFDQDKLVGGAILPGLGLQLATLGDRLAALPKLEMD 167
Db 118 AARVLYPGACIIIVDMGTATTTFTLSEGGAYIGAIAPGIATSAQAIAEKTSLPKIEII 177
QY 168 QLTLEPDRWALDTPSAIFSGVYVGLGALQSYLQ 201
Db 178 RPAKV--IGSNTVSAMQSGIYFGIYGLVEELVR 208
RESULT 15
AAW98422
ID AAW98422 standard; Protein; 223 AA.
XX AC AAW98422;
XX DT 31-MAR-1999 (first entry)
XX DE H. pylori GHPO 344 protein.
XX KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX KW peptic ulcer disease.
XX OS Helicobacter pylori.
XX PN WO9843478-A1.
XX PD 08-OCT-1998.
XX PF 01-APR-1998; 98WO-US06371.
XX PR 29-JUL-1997; 97US-0902615.
XX PR 01-APR-1997; 97US-0833457.
XX PR 24-JUN-1997; 97US-0881227.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX WPI; 1998-542293/46.
XX DR N-PSDB; AAX14141.
XX PT New isolated Helicobacter polynucleotides - used to develop products
XX PT for the diagnosis, prevention and treatment of Helicobacter
XX PT infections and gastrointestinal diseases
XX PS Claim 8; Page 709-710; 2054pp; English.
XX This sequence represents a Helicobacter pylori GHPO protein of the
XX invention. The polypeptides can be used for preventing or treating
XX Helicobacter infections, and gastroduodenal diseases associated with
XX these infections, including acute, chronic, and atrophic gastritis, and
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
XX used for the production of antibodies. The products can also be used for
XX detection and diagnosis.
XX SQ Sequence 223 AA;
Query Match 10.3%; Score 141.5; DB 19; Length 223;

. Best Local Similarity 26.9%; Pred. No. 2.8e-05;
Matches 66; Conservative 28; Mismatches 76; Indels 75; Gaps 12;

QY	25	IGNSRLHWA-----YCSGNAPLQWVTWTDYNPKSAQLPVLGKVPPLMLASVVPPEQTEVWRV	79
.Db	18	IGNTRIHPAQNQLFSSAKEDLKR-----LGIQKEIFYISVNEENE-----	58
QY	80	YQPKILTTLKNLPLVNLPS-----FGIDRALAGLGTGTYGFPCLVVDG	123
Db	59	-----KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA	103
QY	124	GTALTITGPDQDKLVGGAILPGLG--LQATLGRDLAALPKLEMDQLTELPDRWALDTP	181
Db	104	GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAILKPPFKALDSLEVLPK-----STR	158
QY	182	SAIFSGVVYVGLGALQSYLDWQKLPFGAAMVITGGDGKILHGFELKEHSPNLSVAVDDNL	241
Db	159	DAVNYGMVLSVTACIQ-HLAKNQKIY-----LCGGDAKYLSAFL-PHS-----VCKERL	205
QY	242	IFLGM	246
Db	206	VEDGM	210

Search completed: June 24, 2003, 21:46:20
Job time : 26.3857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.37386 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-13

Perfect score: 1368

Sequence: 1 METSKPGCGGLALDNDKQKPW.....DDNLFLGMAAIHHGDRPIC 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1368	100.0	257	S75559	hypothetical prote
2	537	39.3	276	A12292	hypothetical prote
3	174	12.7	265	T36391	hypothetical prote
4	170	12.4	259	AF1102	conserved hypot
5	164	12.0	259	AF1464	conserved hypot
6	158	11.5	233	S66100	conserved hypot
7	147	10.7	261	B87489	transcription acti
8	142.5	10.4	223	G71887	hypothetical prote
9	141.5	10.3	223	F64627	hypothetical prote
10	141	10.3	592	B81009	Bira protein/Bvg a
11	141	10.3	592	H82031	probable biotin-la
12	138	10.1	246	D72320	conserved hypot
13	137.5	10.1	254	F83660	hypothetical prote
14	137.5	10.1	273	E97293	probable transcrip
15	136.5	10.0	273	D71326	conserved hypot
16	133	9.7	229	E70465	hypothetical prote
17	123.5	9.0	262	E75516	conserved hypot
18	121	8.8	274	H86937	conserved hypot
19	116.5	8.5	209	H81382	hypothetical prote
20	107	7.8	248	H83111	hypothetical prote
21	106	7.7	272	A70955	hypothetical prote
22	105	7.7	228	S11803	pullulanase secret
23	100.5	7.3	1070	B86922	probable arabinosy
24	99	7.2	242	A82637	conserved hypot
25	97	7.1	267	I40327	baf protein - Bord
26	95.5	7.0	749	A20345	probable phosphate
27	94.5	6.9	498	H95220	hypothetical prote
28	94.5	6.9	498	S11150	amic protein - Str
29	94.5	6.9	498	H98084	hypothetical prote

hydrogenase access
hypothetical prote
hypothetical prote
polyketide synthas
conserved hypotet
probable amino aci
probable amino aci
probable amino aci
conserved hypotet
hypothetical prote
xylokinase (EC 2
xanthine/uracil pe
probable integral pe
MFS permease (impo
ferrichrome ABC tr
hypothetical prote

ALIGNMENTS

RESULT 1

S75559

hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S75559

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.

S:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S75559

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-257 <KAN>

A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAAL8120.1; PID:d101

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

Query Match

100.0%; Score 1368; DB 2; Length 257;

Best Local Similarity 100.0%; Pred. No. 7.7e-104;

Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METSKPGCGGLALDNDKQKPWLGLMIGNSRLHWAYCSGNAPLQTVWTDYNNPKSAQLPVLIG 60

Db 1 METSKPGCGGLALDNDKQKPWLGLMIGNSRLHWAYCSGNAPLQTVWTDYNNPKSAQLPVLIG 60

QY 61 KVPMLASVVPQTEVWRVYQPKILTILKNLPLVNLVPSFGIDRALAGLGTGLTYGFPCLV 120

Db 61 KVPMLASVVPQTEVWRVYQPKILTILKNLPLVNLVPSFGIDRALAGLGTGLTYGFPCLV 120

QY 121 VDGFTALITGFDQDKLVGGAILPGILGLQALATLGLDRLAALPKLEMDQLTLPDRWALDT 180

Db 121 VDGFTALITGFDQDKLVGGAILPGILGLQALATLGLDRLAALPKLEMDQLTLPDRWALDT 180

QY 181 PSAIFSGVYVYLGALQSYLDQWKLFFGAAAVITGGDGKILHGLKHSPLNSLVAWDDN 240

Db 181 PSAIFSGVYVYLGALQSYLDQWKLFFGAAAVITGGDGKILHGLKHSPLNSLVAWDDN 240

QY 241 LIFLGMAAIHHGDRPIC 257

Db 241 LIFLGMAAIHHGDRPIC 257

RESULT 2

A12292

hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: A12292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena dactyloides
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: A12292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA075595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

Query Match 39.3%; Score 537; DB 2; Length 276;
Best Local Similarity 42.8%; Pred. No. 2.8e-36;
Matches 113; Conservative 43; Mismatches 80; Indels 28; Gaps 3;
17 QKPLWGLMIGNSRLHWAYCSGNAPLOTWTDYDYNPKSA----- 53
11 ENOWALEIGNSRLHWALFGESELEFTWDTYELPESVIQQLGNCETKLEVGSEKEIFFT 70
54 --QLPVLVGLKVPMLASVPEQTEWRY--QPKILTTLKPLVNLVPSFGIDRALAGLT 110
71 FFLPLPAPCPPLPLFIASVPOQTWLNVRVITDQPLNNIYPTLGDRLALWGA 130
111 GLTVGFPCLVVDGTALTITGFDQDKLVGAILPGLGLQATLGDRLAALPKLEMDLT 170
131 GMSGFPVVLVDAAGTALTFTADGCKNLVGGAILPVGGLQFASIGQQTGQLPQVMEAIK 190
171 ELDPDWALDTPSAIFSQVGVGLGALQSYLDQWQKLPFGAAMVITGDKGILHGFLEHS 230
191 SLPPRFALNTTEAIQSGVIITLIAGMRDFTTEWLSLFPDGVKVAIKGDRILLNVLQALY 250
231 PNLS--VAVDDNLIFLGMALHIG 252
251 PDLAARLIVEPNLIFGMQTIAG 274

RESULT 3
T36391
hypothetical protein SCE94.3lc - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T36391
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, April 1999
Reference number: Z1573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.3lc
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOE94.3lc
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match 12.7%; Score 174; DB 2; Length 265;
Best Local Similarity 24.9%; Pred. No. 8.3e-07;
Matches 67; Conservative 45; Mismatches 105; Indels 52; Gaps 11;
21 LGLMIGNSRLHWAYCSGNAPLOTW--VTDYNPKSAQPLVLL-----GKVPML----- 65
3 LTIDVGTHVTLGFDGEDIVEHWRISTDSRRTADELAVLLQGLMGHPILLGDELGDID 62
66 -----LASVVPQTEWRY-----QPKILTTLKPLVNLVPSFGIDRALAGL 108
63 GIAICATVPVSLHRELVRTRYGDPAVLVEPGVKT--GVPILTDPKVEGADRIINAV 120
109 GTGLTYGFPCLVVDGTTALTITGFDQDKLVGAILPGLGLQATLGDRLAALPKLEMDQ 168

Db 121 AAVELYGGPAIVVDGTFATTFFDAVSARGEYIGGVIAFGIEISVBAALGVKAQLRKIEVAR 180
Qy 169 LTELDPDWALDTPSAIFSQVGVGLGALQSYLDQWQKLF---PGAAMVITGGDKILHG 224
Db 181 PRSVIGK---NTVEAMSGVIGYPAGQVGVVNRMARLADDDPDVTVIATGGGLAPVLG 237
Qy 225 ---FLKEHSPNLSVAVDDNLIFLGMALHIG 250
Db 238 ESSVIDSEHP-----W---LTLMGLRLVY 258

RESULT 4

AF1102
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EG)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1102
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, A.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001
A;Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla, A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1102
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0221
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match 12.4%; Score 170; DB 2; Length 259;
Best Local Similarity 27.2%; Pred. No. 1.7e-06;
Matches 56; Conservative 43; Mismatches 83; Indels 24; Gaps 8;
Qy 64 LMLASVYPE-----QTEWVRVYQ-----PKILTTLKPLVNLVPSFGIDRALAGLT 110
Db 60 IITSSVVPPIHMETMCMVRYFNIRPLIVGFKTGLNLKVDN--PREIGSDRIVNVA 117
Qy 111 GLVYGFPCLVVDGTTALTITGFDQDKLVGAILPGLGLQATLGDRLAALPKLEMDLT 170
Db 118 SEEGYFPIVVDGTFATTFCYIDESGVYQGAAPGIMISTEALYNAKLPVVDIAESS 177
Qy 171 ELDPDWALDTPSAIFSQVGVGLGALQSYLDQWQKLPFGAAMVITGDKGILHGFLEH 229
Db 178 QIIGK---STVSSMQAGIFYGVGQCEGIIAEMKQSNQSPVVVATGG----LARMITEK 230
Qy 230 SPNLSVAVDDNLIFLGMALHIGDRP 255
Db 231 SSAVDIL-DPFLTKGLELYRRNK 255

RESULT 5

AF1464
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip1126)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1464
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, A.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001
A;Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla, A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1464
A;Status: preliminary

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: G71897

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-223 <ARN>

A:Cross-references: GB:AE001509; GB:AE001439; NID:g4155350; PIDN:AAD06372.1; PID:g415536

A:Experimental source: strain J99

C:Genetics:

A:Gene: jhp0796

C:Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 10.48; Score 142.5; DB 2; Length 223;

Best Local Similarity 25.94; Pred. No. 0.00025;

Matches 63; Conservative 34; Mismatches 75; Indels 71; Gaps 13;

25 IGNSRLHWA-----YCSGNAPLOTWTDYNNPKSAQLPVLGKVPMLASVVPQTEWVRV 79

18 IGWTRIFHAQNYQLFSSAKEDLKR-----LGIQKEIFYISVNEENE-----58

80 YQPKILTKLPLVNLVPSFGIDRALAG-----LGTGLTYGFPCL-----VVDGGT 125

59 -----KALLNCYPN---AKNIAGFFHLETDYIGLIDRQMACLAVVNGVIVDAGS 105

126 ALRITGFDQDKLVGGAILPGLG--LQATLGDRLAALPKLEMDQLTLPDRWALDTPSA 183

106 AITIDLVEKGKHL--GGCILPGLAQYHAYKKSAILQPPFKALDSLEVLKP-----NURDA 160

184 IFSGVYGVGLGALQSYLDQWQKLPFGAAMVITGGDKILHGLKHSPLNSVAWDDNLIF 243

161 VNYGMILSIISCIQHLAKD-QKIY-----LCGGDAKYLSAFL-PHS-----VCKERLVF 207

244 LGM 246

208 DGM 210

RESULT 9

F64627

hypothetical protein HP0862 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000

C:Accession: F64627

R:Tomb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen-

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

Accession: F64627

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-223 <TON>

A:Cross-references: GB:AE000596; GB:AE000511; NID:g2313982; PIDN:AAD07916.1; PID:g231400

C:Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 10.34; Score 141.5; DB 2; Length 223;

Best Local Similarity 26.94; Pred. No. 0.0003;

Matches 66; Conservative 28; Mismatches 76; Indels 75; Gaps 12;

25 IGNSRLHWA-----YCSGNAPLOTWTDYNNPKSAQLPVLGKVPMLASVVPQTEWVRV 79

18 IGWTRIFHAQNYQLFSSAKEDLKR-----LGIQKEIFYISVNEENE-----58

80 YQPKILTKLPLVNLVPS-----FGIDRALAGLGTGYGFPCLVWDG 123

59 -----KALLNCYPNKNAGTGFHLETDYVIGLIDRQMACLAVN-----NGVVDA 103

124 GTALTITGFDQDKLVGGAILPGLG--LQATLGDRLAALPKLEMDQLTLPDRWALDTP 181

104 GSAITIDLIEGKHL--GGCILPGLAQYIHAYKKSAILQPPFKALDSLEVLKP-----STR 158

182 SAIFSGVYGVGLGALQSYLDQWQKLPFGAAMVITGGDKILHGLKHSPLNSVAWDDNL 241

Db 159 DAVNYGMVLSVIAQIO-HLAKNOKIY-----LCGGDAKYLSAFL-PHS-----VCKERL 205

Qy 242 IFLGM 246

Db 206 VFDGM 210

RESULT 10

B81009

BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: B81009

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

ri, H.; Qin, H.; Vanathavan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: B81009

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-592 <TET>

A:Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB2075

Query Match 10.34; Score 141; DB 2; Length 592;

Best Local Similarity 27.74; Pred. No. 0.001;

Matches 69; Conservative 37; Mismatches 99; Indels 44; Gaps 13;

Qy 26 GNSRLHWAYCS-----GNAPLOTWTDYNNPKSAQLPVLGKVPMLASVVPQTEWVRV 401

Db 346 GNSRLKWAWEVNGTGFVGSAPYR-----DLSPLGAWEAKADGNVRIVGCAVGEFKKAQ 401

Qy 76 VWRVYQPKILTILKN-----LPLVNLV-----PSFGIDRALAGLGTGYGFPCLVVDGGTALT 128

Db 402 VQEQARKIEWLPSSQAQALGIRNHRHPPEHSGDRWFNLGSRFRSNCACVVVSCGTAVT 461

Qy 129 ITGFDQDKLVGGAILPGLGLOLATLGDRLAALPKLEMDQLTLPDRWALDTP--SAIFS 186

Db 462 VDALTDDHGYLGTGIMPFGFLMKESLAVRTANLNR-----HAGKRYPPPTTTGNVAVS 514

Qy 187 GVVYGVGLGALQSYLDQWQKLPFGAA-----VWITGGD-GKTLH-----GFLKEHSPNLSVAV 237

Db 515 GMDAVCGSVW--MMHGRLEKTKGAKRPDVIIITGGGNAKVAEALPPAFLAENTVRVA---570

Qy 238 DDNLIFLGM 246

Db 571 -DNLVIYGL 578

RESULT 11

H82031

probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitidis

N:Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002

C:Accession: H82031

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: H82031

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-592 <PAR>

A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737

A:Experimental source: serogroup A, strain 22491

RESULT 15

D71326
 conserved hypothetical protein TP0431 - syphilis spirochete
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
 C:Accession: D71326
 R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwilt
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-388, 1998
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A:Reference number: A71250; MUID:98332770; PMID:9665876
 A:Accession: D71326
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-273 <COL>
 A:Cross-references: GB:AE001220; GB:AE000520; MID:g3322705; PIDN:AAC65417.1; PID:g332271
 A:Experimental source: strain Nichols
 C:Genetics:
 Gene: TP0431
 Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 10.08; Score 136.5; DB:2; Length 273;
 Best local Similarity 31.18; Pred. No. 0.00095;
 Matches 51; Conservative 25; Mismatches 75; Indels 13; Gaps 6;
 QY 62 VPLMLASVVPETEVRWVYQPKIL---TLKNLPLVNLXP---SFGIDRALAGLTGLTYG 115
 DB 69 VPLTKTIADAVAQISGV-QPVVFGPWAYEHLPRPEPVRAEIGTDLVANAVAAVHFR 127
 QY 116 FPCLVVDGGTALTITGFDQDKLVGAILFGLGLQATLGLDRLAALPKLEMDQLTELPDR 175
 DB 128 SACVYVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTGTAQLPLVPL----ALPDS 183
 QY 176 -WALDTPSAIFSGVYGVIGALQSYLQDWQK-LFPGAAVMVITGG 217
 DB 184 VLGKDTTHAVQGVWRTGLEFVIRAMIAQCCKELGCRCAAVITGG 227

Search completed: June 24, 2003, 22:03:12
 Job time : 10.3739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.46994 Seconds

(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-13

Perfect score: 1368

Sequence: 1 METSKPGGLALDNDKQKPW.....DDNLFLGMAAIHGGDRPIC 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	158	11.5	233	1 YACB_BACSU	P37564 bacillus su
2	105	7.7	228	1 LEP4_KLEPN	P15754 klebsiella
3	100.5	7.3	1070	1 EMB6_MYCLE	Q9cda7 mycobacteri
4	97	7.1	267	1 BAF_BORPE	Q45338 bordetella
5	97	7.1	452	1 Y397_CLOPE	P50487 clostridium
6	96	7.0	481	1 XYL6_STRCP	Q9ck00 streptomyc
7	94.5	6.9	498	1 AMTC_STRPN	P18793 streptococ
8	91.5	6.7	291	1 LEP4_AERSA	O68964 aeromonas s
9	91	6.7	454	1 YACB_BACSU	P42314 bacillus su
10	91	6.7	461	1 YCUJ_ECOLI	P76037 escherichia
11	89.5	6.5	481	1 XYL6_STRRP	P27156 streptomyc
12	87.5	6.4	380	1 PGL1_SCLSC	Q12708 sclerotinia
13	87	6.4	532	1 YABM_BACSU	P37555 bacillus su
14	84.5	6.2	475	1 ALG6_DROME	Q9vax7 drosophila
15	84	6.1	247	1 TPIS_CHICK	P00940 gallus gall
16	84	6.1	369	1 FIXB_BRAJA	P10449 bradyrhizob
17	84	6.1	514	1 AT12_ARATH	Q92pj8 arabidopsis
18	83.5	6.1	405	1 TRPB_PSEPU	P11080 pseudomonas
19	83	6.1	228	1 RNFG_PSEST	Q9evn3 pseudomonas
20	83	6.1	452	1 MLTD_ECOLI	P23931 escherichia
21	83	6.1	1061	1 CPFD_BACSU	O08394 bacillus su
22	82.5	6.0	247	1 TPIS_LATCH	P00941 latimeria c
23	82.5	6.0	360	1 FIXB_AZOVI	P53574 azotobacter
24	82.5	6.0	479	1 TPIS_VIBAL	P22825 vibrio algi
25	82.5	6.0	894	1 YTHL_ECOLI	P37624 escherichia
26	81	5.9	290	1 LEP4_AERHY	P45794 aeromonas h
27	81	5.9	747	1 ATCS_SYNP7	P37279 synecococc
28	81	5.9	2482	1 VNF_PIG	Q28833 sus scrofa
29	80.5	5.9	414	1 SYH_MYCPN	P75069 mycoplasma
30	80.5	5.9	529	1 GUAA_MYCLE	P46810 mycobacteri
31	80.5	5.9	808	1 FGR4_MOUSE	O03142 mus musculu
32	80	5.8	311	1 SUFI_HAEIN	P44847 haemophilus
33	80	5.8	470	1 PTIB_HACIN	P39794 bacillus su

RESULT 1

ID	YACB_BACSU	STANDARD;	PRT;	233 AA.
AC	P37564;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical protein yacB.			
GN	YACB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=96051385; PubMed=7584024;			
RA	Ogasawara N., Nakai S., Yoshikawa H.;			
RT	"Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chromosome containing the replication origin.";			
RL	DNA Res. 1:1-14(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Furnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";			
RL	Nature 390:249-256(1997).			
CC	!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.			

ALIGNMENTS

34	80	5.8	613	1 NUOL_ECOLI	P33607 escherichia
35	80	5.8	638	1 S6A9_BOVIN	Q28039 bos taurus
36	79.5	5.8	614	1 VAA1_DROME	P48602 drosophila
37	79.5	5.8	849	1 Y343_METJA	Q57789 methanococc
38	79	5.8	462	1 IL6A_RAT	P22273 rattus norv
39	79	5.8	1094	1 EMB6_MYCTU	P72059 mycobacteri
40	79	5.8	2076	1 FAS1_YARLI	Q47229 y fatty aci
41	78.5	5.7	547	1 SAPA_ECOLI	Q47622 escherichia
42	78.5	5.7	633	1 NUOL_MYCTU	O86350 mycobacteri
43	78.5	5.7	2051	1 FAS1_YEAST	P07149 s fatty aci
44	78	5.7	258	1 DDAH_STRCO	Q9X7M4 streptomyc
45	77.5	5.7	407	1 YM94_MYCTU	Q50672 mycobacteri


```

122 PLLVASFQATTTLDITGPDNVFPGGLILGGPAMRGALAYGTATHLPAD-----GIIVAD 176
176 WALDTPSAIFSGVVYGVGLAGL-OSYLQDWKQLFPFGAAMVITGG 217
      : :: || ||| : : : : :
177 YPIDTHQTAIASGIAAAQAAGAIVRQWLGRQRVYGAPEIYVAGG 219

RESULT 5
Y397_CLOPE
ID Y397_CLOPE STANDARD; PRT; 452 AA.
AC P50487; P94653;
DT 01-OCT-1996 (Rel. 34, Created)
DD 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DEF Putative purine permease CPE0397.
GN CPX OR CPE0397.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
CC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RA PubMed=11792842;
RX Shimizu T., Ohnami K., Hirakawa H., Ohsheima K., Yamashita A.
```



```
DB 357 QAAVDGAVHSLGALDLVLD--ADADPSAPILLIGGAR-----GTAWQOTV 401
QY 240 -----NLIFLGMAA 248
      : : : :
DB 402 RRLSGRPVQIPEARELVALGAA 424

RESULT 7
AMIC_STRPN
ID AMIC_STRPN STANDARD; PRT; 498 AA.
AC P18793;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Oligopeptide transport system permease protein amic.
OS AMIC OR SP1890.
GN Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1313;
RP [1]
SEQUENCE FROM N.A.
RC STRAIN=R800;
RX MEDLINE=90279506; PubMed=2352474;
RA Allouing G., Trombe M.C., Claverys J.-P.;
RT "The ami locus of the Gram-positive bacterium Streptococcus pneumoniae
RT is similar to binding protein-dependent transport operons of Gram-
RT negative bacteria.";
RL Mol. Microbiol. 4:633-644(1990).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR OLIGOPEPTIDES; PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF
CC THE SUBSTRATE ACROSS THE MEMBRANE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. OPPEC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17337; CAA35214.1; -.
CC DR EMBL; AE007479; AAK75961.1; -.
CC DR PIR; S11150; S11150.
CC DR TIGR; SP1890; -.
CC DR Pfan; PF00528; BPD_transp; 1.
CC DR PROSITE; PS00402; BPD_TRANS_PNN_MEMBR; FALSE_NEG.
KW Transport; Peptide transport; Transmembrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 279 299 POTENTIAL.
FT TRANSMEM 316 336 POTENTIAL.
FT TRANSMEM 359 379 POTENTIAL.
FT TRANSMEM 415 435 POTENTIAL.
FT TRANSMEM 461 481 POTENTIAL.
```

```
SQ SEQUENCE 498 AA; 55623 MW; B9B30A794C3B52DC CRC64;
Query Match 6.9%; Score 94.5; DB 1; Length 498;
Best Local Similarity 23.6%; Pred. No. 1.6;
Matches 59; Conservative 29; Mismatches 85; Indels 77; Gaps 13;
QY 10 LALDNKOKPWLGLMIGNSRLHWAYCSGNAPL--QTWYT----DYNPKSAQLPVLLGRVP 63
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 160 LRFPNDPAIGNSLVGSGTKHKVLLYFNSQFPVHQNFVNLNLGDSYPTVANTPVL----- 214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 LMLASVVEP---QTEVWRVYQPKILTILKNLPLVNLV-----PSGIDRALAGLG----- 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 215 ----QVITQGGQGTAKTAQVFP---TGKTKSSVNIYSRTYKSPQADRSVASYGKDDPY 267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 --TGLTYGFCPLVVDGTTALITGFDQDKKLKVGGAIPGLGLQIATLIGDR----- 157
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 268 TATESNYQYPSMIWSS-----AITG-----LIGLVAYALAVPLGSAMARFNTWIDSLS 317
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 -----LAALPKLEMDQLTE-----LPDRWAL-----DTPSAIFSGVYVGLA----- 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 318 TGAITFLALPTIALVIVRLIGSIALPDSFPILGAGDWSRYVLPVAVILGLGAPGTAI 377
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 196 -LQSYLDQWQ 204
      : : : : :
DB 378 WIRRYMIDLQ 387

RESULT 8
LEP4_AERSA
ID LEP4_AERSA STANDARD; PRT; 291 AA.
AC O68964; O54483;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Type 4 prepilin-like proteins leader peptide processing enzyme
DE [includes: Leader peptidase (EC 3.4.99.-) (Prepilin peptidase); N-
DE methyltransferase (EC 2.1.1.-)].
GN TAPD OR PILD.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=645;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=A450;
RA Pepe C.M., Strom M.S.;
RT "Aeromonas salmonicida type IV prepilin peptidase and type IV pilus
RT assembly genes.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=A449;
RA Lutwyche P., Perez-Casal J.F., Crump E.M., Trust T.J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CLEAVES TYPE-4 FIMBRIAL LEADER SEQUENCE AND METHYLATES
CC THE N-TERMINAL (GENERALLY PHE) RESIDUE. PROCESSES THE TAPA PILIN
CC PRECURSOR DURING MEMBRANE TRANSLLOCATION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF059249; AAC23569.1; -.
CC DR EMBL; U95640; AAC04561.1; -.
CC DR MEROPS; A24.001; -.
CC DR InterPro; IPR000045; Peptidase_C20.
```

DR Pfam; PF01478; Peptidase_C20; 1.
 DR PRINTS; PRO0864; PREPLINFRASE.
 KW Multifunctional enzyme; Hydrolase; Protease; Transferase;
 KW Methyltransferase; Transmembrane; Inner membrane.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 262 282 POTENTIAL.
 FT CONFLICT 60 80 P -> A (IN REF. 2).
 FT CONFLICT 84 84 A -> G (IN REF. 2).
 FT CONFLICT 100 100 C -> W (IN REF. 2).
 FT CONFLICT 103 103 C -> W (IN REF. 2).
 FT CONFLICT 126 126 A -> G (IN REF. 2).
 FT CONFLICT 163 163 T -> N (IN REF. 2).
 FT CONFLICT 198 198 S -> R (IN REF. 2).
 FT CONFLICT 253 253 L -> V (IN REF. 2).
 SQ SEQUENCE 291 AA; 32500 MW; 4CFFA0AFC886E368 CRC64;
 Query Match 6.7%; Score 91.5; DB 1; Length 291;
 Best Local Similarity 21.08; Pred. No. 1.5;
 Matches 62; Conservative 45; Mismatches 87; Indels 101; Gaps 13;
 QY 19 PWL-----GLMIGNSRLHWAYCSGNAPLOTWTDYDYNPKSAQLPVLGKVPMLA--- 67
 DB 12 PLYFSLVFLSLMISGF-----LNVIHRLPIMLEREW 45
 QY 68 -----SVVPEQTEWVRVYQPK-----ILTLKLNPLVNLVPSFGIDRALAG 107
 DB 46 QAEYRSVFSSTDPQEDDERYNLWVRSCPCPRNHPITALENTPLLSWLKRCR---G 102
 QY 108 LGTGLTYGFCPLVVDGTLTIGFDODKLVGGAILPGLG-LQATLGDRLAALPKLEM 166
 DB 103 CQRAISRYP-LVELLTALL-----SVVAMTLTPGWGTLAALLTWLVALLFIDL 153
 QY 167 DQTELPDRWALTPSAIFSGVYV-----GYLGALOSYLODQWKLFPGAAMVIT 215
 DB 154 DKML-LPDQLTL---PLLWGLLFLNLGGVYVPLGDAVIGAGYLVLSLYW--AFKLLT 207
 QY 216 GGGKILHGLKHESPNLSVANDNLIFLGMAI-----HHGDRPI 256
 DB 208 GKEGMVGDFPKLLAALGAWLQWALPIVLLSLVGAIFGIGLILRNHHQSKPI 262
 RESULT 9
 YXJC_BACSU STANDARD; PRT; 454 AA.
 P42314;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein yxjC.
 GN YXJC OR N15J.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=97124196; PubMed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miwa Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the lic and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacXy region.";
 RL Microbiology 142:3113-3123(1996).
 -RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V.; Bertero M.G.; Bessieres P.; Bolotin A.; Borchert S.,
 RA Brouillet R.; Boursier L.; Brans A.; Braun M.; Brignell S.C.; Bron S.,
 RA Choi S.K.; Codani J.J.; Connerion I.F.; Cummings N.J.; Daniel R.A.,
 RA Denizot F.; Devine K.M.; Dusterhoft A.; Ehrlich S.D.; Emmerson P.T.,
 RA Entian K.D.; Errington J.; Fabret C.; Ferrari E.; Foulger D.,
 RA Fritz C.; Fujita M.; Fujita Y.; Fuma S.; Gallizzi A.; Galleron N.,
 RA Ghim S.Y.; Glaser P.; Goffeau A.; Golightly E.J.; Grandi G.,
 RA Guiseppe G.; Guy B.J.; Haga K.; Halech J.; Harwood C.R.; Henaut A.,
 RA Hilbert H.; Holsappel S.; Hosono S.; Hullo M.F.; Itaya M.; Jones L.,
 RA Joris B.; Karamata D.; Kasahara Y.; Kjaer-Blanchard M.; Klein C.,
 RA Kobayashi Y.; Koetter P.; Koningsstein G.; Krogh S.; Kumano M.,
 RA Kurita K.; Lapidus A.; Lardinois S.; Lauber J.; Lazarevic V.,
 RA Lee S.M.; Levine A.; Liu H.; Masuda S.; Mauel C.; Medigue C.,
 RA Medina N.; Mellado R.P.; Mizuno M.; Moestl D.; Nakai S.; Noback M.,
 RA Noone D.; O'Reilly M.; Ogawa K.; Ogiwara A.; Oudega B.; Park S.H.,
 RA Parro V.; Pohl T.M.; Portetelle D.; Porwollik S.; Prescott A.M.,
 RA Presecan E.; Pujic P.; Purnelle B.; Rapoport G.; Rey M.; Reynolds S.,
 RA Rieger M.; Rivolta C.; Rocha E.; Roche B.; Rose M.; Sadaie Y.,
 RA Sato T.; Scanlan E.; Schleich S.; Schroeter R.; Scoffone F.,
 RA Sekiguchi J.; Sekowska A.; Seror S.J.; Serror P.; Shin B.S.; Soldo B.,
 RA Sorokin A.; Tacconi E.; Takagi T.; Takahashi H.; Takemaru K.,
 RA Takeuchi M.; Tanakoshi A.; Tanaka T.; Terpstra P.; Tognoni A.,
 RA Tosato V.; Uchiyama S.; Vandenberg M.; Vannier F.; Vassarotti A.,
 RA Viari A.; Wambutt R.; Wedler E.; Wedler H.; Weitzenecker T.,
 RA Winters P.; Wipat A.; Yamamoto H.; Yamane K.; Yasumoto K.; Yata K.,
 RA Yoshida K.; Yoshikawa H.F.; Zumstein E.; Yoshikawa H.; Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D83026; BAAL1704.1; -;
 DR EMBL; Z99123; CAB15926.1; -;
 DR EMBL; Z99124; CAB15936.1; -;
 DR Subtilist; BG1152; yxjC.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 56 76 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 140 160 POTENTIAL.
 FT TRANSMEM 176 196 POTENTIAL.
 FT TRANSMEM 240 260 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT TRANSMEM 372 392 POTENTIAL.
 SQ SEQUENCE 454 AA; 48293 MW; FF93F4EE52D11CA CRC64;
 Query Match 6.7%; Score 91; DB 1; Length 454;
 Best Local Similarity 20.38; Pred. No. 2.8;
 Matches 61; Conservative 34; Mismatches 91; Indels 114; Gaps 13;
 QY 19 PWLGLMIG-----NSRLHWAYCSG-----NAPL-OTWTDYDYNPKSAQLPV 57
 DB 180 PWLGLMIGAVIVLAAGMAYLESRRKKAQASGEYGFDSQNAPESESIAEAPDKSPIRH 239
 QY 58 LLGKVPMLASVVPTEQTEWVRVYQPKILKLNPLVNLVPSFGIDR----- 103
 DB 240 ALAFVPLILVGV---NKYFTIYLPKW-----YPNGFNFPISIGLKEFRGLDISSAAIWS 291
 QY 104 -----ALAGLGTGTYGFCPLVVDGTLT 128
 DB 292 VEIALVIGIITITLFDWRSVFAQLKEGLNEGGALLASMTGAEYGF-----GGIIAA 345

FT	TRANSMEM	54	74	POTENTIAL.
----	----------	----	----	------------

PFam; PF02782; FGGY_C; 1.
TIGRFAMs; TIGR01312; xylB; 1.

FT	TRANSMEM	54	74	POTENTIAL.
----	----------	----	----	------------


```

RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Fujic C., Furnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., SEKOWSKA A., Seror S.J., Sarrot P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vanmier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenegger T.,
RA Winter P., Wipat A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshioka H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT *The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".;
RT Nature 390:249-256(1997).
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC CC -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE SYNTHASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or
CC send an email to licensese@isb-sib.ch)
-----
CC EMBL; D26185; BAA05292.1; .
CC EMBL; Z99104; CAB11833.1; -.
CC Subtilist; BG10120; yabM.
CC InterPro; IPR002797; Polysacc_synt.
CC Pfam; PF01943; Polysacc_synt_1.
KW Hypothetical Protein; Transmembrane; Complete proteome.
FT TRANSMEM 25      45   POTENTIAL.
FT TRANSMEM 65      85   POTENTIAL.
FT TRANSMEM 109     129   POTENTIAL.
FT TRANSMEM 134     154   POTENTIAL.
FT TRANSMEM 179     199   POTENTIAL.
FT TRANSMEM 203     223   POTENTIAL.
FT TRANSMEM 248     268   POTENTIAL.
FT TRANSMEM 302     322   POTENTIAL.
FT TRANSMEM 344     364   POTENTIAL.
FT TRANSMEM 371     391   POTENTIAL.
FT TRANSMEM 392     412   POTENTIAL.
FT TRANSMEM 425     445   POTENTIAL.
FT TRANSMEM 459     479   POTENTIAL.
FT TRANSMEM 494     514   POTENTIAL.
SQ SEQUENCE       532 AA; 57357 MW; DAACAFAF32903E553 CRC64;
Query Match          6.4%; Score 87; DB 1; Length 532;
Best Local Similarity 24.5%; Pred. No. 7.1;
Matches 51; Conservative 33; Mismatches 84; Indels 40; Gaps 10;
QY    19 PWLGIMIGNSRLHWAYCSGNAPLTQTWTDYNPKSAOLPVLLGKVPVLMLASVYPQEOTEVWR 78
Db    303 PYISMAVNKNKEIKMKKETITSSLCIVLGTGSAGLICILKPVTNLMIFQ-NGETGLAQ 361
QY    79 VTQPKILLTKNLPLNVLPSPFGIDRALAGLGTGLATYGFPCLVVDDGGTALTITTFDQDKKL 138
Db    362 VFESCI-----LFPSALVTAADVLOGAGYT-V-FPAIAVGAGVAV-----KKW 402
QY    139 VGGAIIPLGLQLATLGD--RLAAPLKLEMDQITELPDRWDLPSPAIFSGVVGYLVGA- 195
Db    403 LNTLIIVPYRGIEGASLATASAFAVAAGLNLYQLRQ-KEW-LDK----LRGVLIPTIGSA 455
QY    196 -----LSQYLDMOKIPLP-----GAAMV 213
Db    456 LLMSAVLLATYRLTWTLFFPATGRGRAVI 483
RESULT 14
ID ALG6_DROME STANDARD; PRG; 475 AA.
AC Q9VRX7; Q960z1;
DT 15-JUN-2002 (Rel. 41, Created)
```

15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable dolichyl pyrophosphate Man9GlcNac2 alpha-1,3-
DE glucosyltransferase (EC 2.4.1.-) (Dolichyl-p-Glc:Man9GlcNac2-pp-
DE dolichyl glucosyltransferase).
GN CG5091.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobaraj C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RN Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champ M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Paclab J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Adds the first glucose residue to the lipid-linked
CC oligosaccharide precursor for N-linked glycosylation. Transfers
CC glucose from dolichyl phosphate glucose (dol-p-Glc) onto the
CC lipid-linked oligosaccharide Man(9)GlcNac(2)-pp-Dol (By
CC similarity).
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum (Potential).
CC -!- SIMILARITY: BELONGS TO THE ALG6/ALG8 GLUCOSYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL: AF0031628; AAF52930.1; -.
EMBL: AY01761; AAK93185.1; -.
FlyBase: FBgn0032234; CG5091.
InterPro: IPR004856; Alg6_Alg8.
Pfam: PF03155; Alg6_Alg8; 1.
Hypothetical protein: Transferase; Glycosyltransferase; Transmembrane;
Endoplasmic reticulum.
TRANSMEM 114 133 POTENTIAL.
TRANSMEM 161 181 POTENTIAL.
TRANSMEM 235 255 POTENTIAL.
TRANSMEM 296 316 POTENTIAL.
TRANSMEM 322 342 POTENTIAL.
TRANSMEM 385 405 POTENTIAL.
TRANSMEM 418 438 POTENTIAL.
TRANSMEM 441 461 POTENTIAL.
CONFLICT 278 292 WCANVVMVKLKHIS -> C (IN REF. 1).
SEQUENCE 475 AA; 53652 MW; C9F67AE977E30E60 CRC64;

RESULT 15

ID	TPIS_CHICK	STANDARD;	PRT;	247 AA.
AC	P00940;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-NOV-1988 (Rel. 09, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Triosephosphate isomerase (EC 5.3.1.1) (TIM).			
GN	TPil.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OC	NCBI_TaxID=9031;			
OC	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=85166263; PubMed=3885220;			
RA	Straus D., Gilbert W.			
RT	"Chicken triosephosphate isomerase complements an <i>Escherichia coli</i>			
RL	deficiency.";			
RL	proc. Natl. Acad. Sci. U.S.A. 82:2014-2018(1985).			
RP	[2]			
RP	SEQUENCE FROM N.A.			

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 20.961 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-13
Perfect score: 1368
Sequence: 1 METSKPGGLALNDNDKQKPW.....DDNLIFLGMAAIIHGGDRPIC 257

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1368	100.0	257	16 P74045	P74045 synechocyst
2	537	39.3	276	16 Q8YQD7	Q8YQD7 anabaena sp
3	175	12.8	295	16 Q8Y2M4	Q8Y2M4 raistonia s
4	174	12.7	265	16 Q9X8N6	Q9X8N6 streptomyce
5	170	12.4	259	16 Q8YAC5	Q8YAC5 listeria mo
6	166	12.1	255	16 Q8R7M2	Q8R7M2 thermoanaer
7	164	12.0	259	16 Q92F54	Q92F54 listeria in
8	152	11.1	258	2 Q9F985	Q9F985 bacillus st
9	147	10.7	261	16 Q9A6Z1	Q9A6Z1 caulobacter
10	142.5	10.4	223	16 Q9ZKY6	Q9ZKY6 helicobacte
11	141.5	10.3	223	16 Q25533	Q25533 helicobacte
12	141	10.3	592	16 Q9JXF1	Q9JXF1 neisseria m
13	141	10.3	592	16 Q9JW17	Q9JW17 neisseria m
14	138	10.1	246	16 Q9WZ15	Q9WZ15 thermotoga
15	137.5	10.1	254	16 Q9RGH5	Q9RGH5 bacillus ha
16	137.5	10.1	273	16 Q97EB4	Q97EB4 clostridium

17	136.5	10.0	273	16	083446	083446 treponema p
18	134	9.8	256	16	08RFEA	08RFE4 fusobacteri
19	133	9.7	229	16	067753	067753 aquifex aeo
20	126.5	9.2	259	16	08XHL5	08XHL5 clostridium
21	123.5	9.0	262	16	09RXS4	09RXS4 deinococcus
22	121	8.8	274	16	09CD56	09CD56 mycobacteri
23	116.5	8.5	209	16	09PIA9	09PIA9 campylobact
24	107	7.8	248	16	09HWC1	09HWC1 pseudomonas
25	106	7.7	272	16	06G282	06G282 mycobacteri
26	103.5	7.6	1046	2	068441	068441 agrobacteri
27	99	7.2	242	16	09PC14	09PC14 xylella fas
28	98	7.2	502	16	098B85	098B85 rhizobium l
29	97	7.1	524	2	093HF4	093HF4 streptomyce
30	95.5	7.0	749	16	08ZCX3	08ZCX3 versinia pe
31	95	6.9	480	2	09SEF5	09SEF5 streptomyce
32	95	6.9	3322	16	08XQ25	08XQ25 raistonia s
33	94.5	6.9	294	16	092KU0	092KU0 rhizobium m
34	94	6.9	3023	12	081487	081487 hepatitis c
35	92.5	6.8	212	2	032514	032514 desulfovibr
36	92.5	6.8	258	16	P73671	P73671 synechocyst
37	92.5	6.8	871	16	08YMP2	08YMP2 anabaena sp
38	92	6.7	1531	10	09LXW7	09LXW7 arbidopsi
39	91.5	6.7	8563	2	Q54297	Q54297 streptomyce
40	91	6.7	461	16	08X7F9	08X7F9 escherichia
41	90	6.6	262	16	051477	051477 borrelia bu
42	90	6.6	384	17	08ZT41	08ZT41 pyrobaculum
43	90	6.6	478	16	09L109	09L109 streptomyce
44	90	6.6	874	2	Q9ZFH0	Q9ZFH0 azotobacter
45	39.5	6.5	241	16	Q9A788	Q9A788 caulobacter

ALIGNMENTS

RESULT 1

P74045
ID P74045 PRELIMINARY; PRT; 257 AA.
AC P74045;
DT 01-PE3-1997 (Tremblrel. 02, Created)
DT 01-PE3-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein slr0812.
GN SLR0812.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyaajima N., Hirotsawa M., Sugliura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimp S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90911; BAA18120.1; -;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 257 AA; 27702 MW; 4291E51EE91FB4A2 CRC64;

Query Match 100.0%; Score 1368; DB 16; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.9e-108;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METSKPGGLALNDNDKQKPWLGLMIGNSRLHWAYCSGNAPLOTWTWTDNPKSAQLPVLIG 60
|||||
DB 1 METSKPGGLALNDNDKQKPWLGLMIGNSRLHWAYCSGNAPLOTWTWTDNPKSAQLPVLIG 60
|||||

```
Qy 61 KVPMLASVPEQTEVVRVYQPKILTLKLNPLVNLVPSFGIDRALAGLGTGLVGFPCLV 120
Db 61 KVPMLASVPEQTEVVRVYQPKILTLKLNPLVNLVPSFGIDRALAGLGTGLVGFPCLV 120
Qy 121 VDGSTALTITGFDQDKKLVGAILPGLGLQATLGLDRLAALPKLEMDQLTELPDRWALDT 180
Db 121 VDGSTALTITGFDQDKKLVGAILPGLGLQATLGLDRLAALPKLEMDQLTELPDRWALDT 180
Qy 181 PSALFSGVYGVGLGALQSLQDMQKLPFGAAMVITGGDKILHGFLEKHSFNLSVAMDDN 240
Db 181 PSALFSGVYGVGLGALQSLQDMQKLPFGAAMVITGGDKILHGFLEKHSFNLSVAMDDN 240
Qy 241 LIFLGMAAIIHGDPRIC 257
Db 241 LIFLGMAAIIHGDPRIC 257

RESULT 2
Q8YQD7 PRELIMINARY; PRT; 276 AA.
ID AC Q8YQD7
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Putative transcription regulation accessory factor transcription
DE regulator protein.
GN RSC0311 OR RS03278.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646058; CAD13839.1;
DR InterPro; IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
KW Complete proteome.
SQ SEQUENCE 276 AA; 30272 MW; 2AD04CA693E56F25 CRC64;

Query Match 39.3%; Score 537; DB 16; Length 276;
Best Local Similarity 42.8%; Pred. No. 9.4e-38;
Matches 113; Conservative 43; Mismatches 80; Indels 28; Gaps 3;

17 QKFWLGMIGNSLRWAYCSGNAPLQTVWTDYDYNPKSA-----
11 ENQWLALEIGNSLRWALFWGESLEFTWDTFYLPEVSIQQLNGCTKLEVGSEKEIFFT 70
54 --QLPVLGLKVPMLASVPEQTEVVRVY--QPKILTLKLNPLVNLVPSFGIDRALAGLGT 110
71 FFPLPPAPCPPLPLFIASVWPQQTVLWENLVNVRVITLDOPLNNIYPTLIGIDRALALWGA 130
111 GLTYGFFCLVVDGSTALTITGFDQDKKLVGAILPGLGLQATLGLDRLAALPKLEMDQLT 170
131 GMSGFPVLVIDAGLTAFTFAADGKNLVGAILPGLGLQATLGLDRLAALPKLEMDQLT 190
171 ELDPDRWALDTPSALFSGVYGVGLGALQSLQDMQKLPFGAAMVITGGDKILHGFLEKHS 230
191 SLPPRPALNTEAIQSGVITVLIAGMRDFTFEEWLSLFPDGVKVAIKGDRILLNLYQALY 250
231 PNLIS--VAVDDNLIIFLGMAAIIHG 252
251 PDLAARLIVEPNIWFWMQTVAG 274
```

RESULT 3
Q8Y2M4

```
ID AC Q8Y2M4 PRELIMINARY; PRT; 295 AA.
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative transcription regulation accessory factor transcription
DE regulator protein.
GN RSC0311 OR RS03278.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646058; CAD13839.1;
DR InterPro; IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
KW Complete proteome.
SQ SEQUENCE 295 AA; 30990 MW; 133074A7764BFCBB CRC64;

Query Match 12.8%; Score 175; DB 16; Length 295;
Best Local Similarity 25.3%; Pred. No. 6.1e-07;
Matches 75; Conservative
Qy 17 QKFWLGMIGNSLRWAYCSGNAPLQTVWTDYDYNPKSAQLPVLGLKVPMLASVVP----- 71
Db 10 ERPLLLIDAGNTRIKWATAADV-----APPVAPGPTWQHAGARPHDOLA 56
Qy 72 EQTEVVR-----VYQPKILTLKLNPLVNLVPSFGI-----DRALAG-- 107
Db 57 ELVEDWRDCHAGAGMAPPDVWISVAGPALRDALCARIVFDGARLRIVASEAAAGLR 116
Qy 108 -----LGT-----GLTYGFP-----CLVVDGSGTALTITGFDQDKKLVGAILPGLGL 149
Db 117 NGYRDPALGTDRWVGAVGARHAWPDTALLLVTAGTATLTDIVAPDGRFAGGLLPGLTL 176
Qy 150 QLATLGLDRLAALPKLEMDQLTELPD-----RWALDTPSALFSGVYGVGLGALQSLQ 201
Db 177 MMRALGRNTAQLPEIDIGYLAARDDAQAPADVPFSDADNTQDATALGCVTAQAGAI---AQ 233
Qy 202 DWQKL---FPGA-AMVITGGDKILHGFLEKHSFNLSVAMDDNLIIFLGMAAIIHG 251
Db 234 TWQALQAQYPCPYRCVLSGGARAL-----APHLRMPFQMDHNLVLGLQVLAH 282

RESULT 4
Q9X8N6 PRELIMINARY; PRT; 265 AA.
ID AC Q9X8N6
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
```



```

RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RA  Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL  Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2);
RA  Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA  Kinashi H., Hopwood D.A.;
RA  "A set of ordered cosmids and a detailed genetic and physical map for
RT  the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL  Mol. Microbiol. 21:77-96(1996).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A3(2) / M145;
RA  Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA  Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA  Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA  Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA  Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA  Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA  Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA  Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA  Hopwood D.A.;
RT  "Complete genome sequence of the model actinomycete Streptomyces
RT  coelicolor A3(2).";
RL  Nature 417:141-147(2002).
DR  EMBL; AL049628; CAB40880.1; -.
DR  InterPro; IPR004619; Baf.
DR  Pfam; PF03309; Bvg_acc_factor; 1.
DR  TIGRFAMs; TIGR00671; baf; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 12.7%; Score 174; DB 16; Length 265;
Best Local Similarity 24.9%; Pred. No. 6.5e-07;
Matches 67; Conservative 45; Mismatches 105; Indels 52; Gaps 11;

QY 21 LGLMIGNSRLHWACSGNAPLQW--VTDYKPSNQLPVLL-----GKVPML----- 65
DB 3 LTIDVGNTHTVLGLFDGEDIVHEHWRISTDSRSTADELAVLLQGLGMHPLGLDGLGDID 62
QY 66 -----LASVVEPQTEWVRVY-----QPKLTLLKNLPLVNLYP-SFGIDRALAGL 108
DB 63 GIAICATVPSVLHRELVTRYRGDVPVAVPEGVKT--GVPILTDHPKEVGADRIINAV 120
QY 109 GTGLTYGPPCLVVDGGTALTITGFDQDKKLVGGAILPGLGLQLATLGDRLAALPKLEMDQ 168
DB 121 AAVELYGGPAIVDFGTATTEDAVSARGEYIGGVIAPIGIEISVEALGVKGALRKIEVAR 180
QY 169 LTELDPDRWALDTPSAIFSGVYGVGLQSLYQDWKLF----PGAAWVTGGDGKILHG 224
DB 191 PRSVIGK---NIVEAMOSGIVYGAGQVDGVVNRMARLADDPDVTVIATGGLPMPVLG 237
QY 225 ---FLKHSPLNSVAVDDNLIFLGMAAIH 250
DB 238 ESSVIDEHEP-----W-----LTLMLGLRVY 258

RESULT 5
QYAC5
ID Q8VAC5 PRELIMINARY; PRT; 259 AA.
AC Q8VAC5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.

```

```

OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquer F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones I.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1; -.
DR MEROPS; M41.009; -.
DR ListLiSt; LMO00221; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 12.4%; Score 170; DB 16; Length 259;
Best Local Similarity 27.2%; Pred. No. 1.4e-06;
Matches 56; Conservative 43; Mismatches 83; Indels 24; Gaps 8;

QY 64 LMLASVPE-----QTEVVRVYQ-----PKILTLLKNLPLVNLYP-SFGIDRALAGLGT 110
DB 60 ILLISVVPPIHMETMCMVRVFNIRPLVPGCIKTLNKLVDN--PREIGSDRIYNVAAA 117
QY 111 GLTYGPPCLVVDGGTALTITGFDQDKKLVGGAILPGLGLQLATLGDRLAALPKLEMDQ 170
DB 118 SEEGTPTVIVDFGTATTFCYIDESGVYGGVGGAIAPGIMISTEALYNRAALPRVDIAESS 177
QY 171 ELDPDRWALDTPSAIFSGVYGVGLQSLYQDWKLFPGAAWVI-TGGDGKILHFLKEH 229
DB 178 QIIGK---STVSSMQAGIFGVFGQCEGIIAEMKKQSNASPVVATGG----LARMITEK 230
QY 230 SPNLSVAVDDNLIFLGMAAIHGDRP 255
DB 231 SSAVDIL-DPEFLTGLGLLELYRNKP 255

RESULT 6
QYAC5
ID Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
GN TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacterales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013180; AAM25520.1; -.
KW Complete proteome; Hypothetical protein.

```


RESULT 11

Q25533	PRELIMINARY; PRT; 223 AA.
ID	Q25533
AC	Q25533;
DT	01-JAN-1998 (TReMBLrel. 05, Created)
DT	01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE	Hypothetical protein HP0862.
DE	GN H0862.
OS	Helicobacter pylori (Campylobacter pylori).
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC	Helicobacter.
OX	NCBI_TaxID=210;
RN	[1]
RP	SEQUENCE FROM N A.

RA MEDLINE=9/394467; PubMed=252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karpi P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RA "The complete genome sequence of the gastric pathogen *Helicobacter*
RA *pylori*,"
RT Nature 388:539-547(1997).
RL Nature 388:539-547(1997).
DR ENBL: AE000596; AAD07916.1; -.
DR TIGR: HP0862; -.
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRfams: TIGR00671; baf; 1.
KW Hypoetical protein; Complete proteome.
SO SEQUENCE. 223 AA; 24655 MW; C1D13E8BCE1EB9AF CRC64;

DB	59	-----KALLNCYPNARNIAGPFHLETDYVGLGIDRQMACLAVN-----NGVVVDA	103
QY	124	GTALTITGFDOOKKLVGGAIPGLG--LQLATGLDRLAALPKLEMDQLTELPDRWALDTP	181

```

Db 104 GSAITDLIKRKHLL-GCCILPLGLAQYIHAYKKSATILEQPFKALDSLEVLPK-----STR 158
QY 182 SAIFSGVYGVGLGALQVLDWQKLPFGAAWITGGDKILHGLFKHSPNLSVWDDNL 241
Db 159 DAVNYGWLVIACIQ-HLAKNQIY-----ICGGDAKYLSAFL-PHS-----VCKERL 205
QY 242 IFGLM 246
Db 206 VFDGM 210

RESULT 12
Q9JXF1
ID AC Q9JXF1 PRELIMINARY; PRT; 592 AA.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE BirA protein/Bvg accessory factor.
GN NMB2075.
SS Neisseria meningitidis (serogroup B).
SC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B; PubMed=10710307;
RX MEDLINE=20175755;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Hart D.H., Cleberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecio A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002557; AAP42394.1; -.
DR HSSP; P06709; 1BIA.
DR TIGR; NMB2075; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BirA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03039; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00121; BirA_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 592 AA; 64701 MW; 9E27421DA2B41DE3 CRC64;

Query Match 10.3%; Score 141; DB 16; Length 592;
Best Local Similarity 27.7%; Pred. No. 0.0012;
Matches 69; Conservative 37; Mismatches 99; Indels 44; Gaps 13;

QY 26 GNSRLHWAYCS-----GNAPLQWTWTDYNPKSAQLPVL-GKVPMLASVYPE--QTE 75
Db 346 GNSRLKWAVENTGTATVGSAPYR-----DLSPLGAWEAKADGNVRVGCVCGEFKAQ 401

QY 76 VWRVYQPKILTAKN-----LPLVNLV-----PSFGIDRALAGLGTGLTYGFCPLVVDGGTALT 128
Db 402 VOELARKIEWLPSSAQAALGIRNHYRHPPEHSGSDRWENALGSRFRSENACVVVSCGTAVT 461

QY 129 ITGFDQDKLVGGAIPGLGLOLATLGDRLAALPKLEMDQLTLPDRWALDTP--SAIFS 186
Db 462 VDALTDDGHVLTGTPGPFHLMKESLAVRTANLNR-----HAGKRYPPPTTTGNVAVS 514

QY 187 GVVYGVGLGALQSLQDWQKLPFGAA-----WVITGGD-GKILH-----GFLKEHSPNLSVAV 237
Db 515 GMDAVCGSVN--MMHGRLEKTKGAGKVPDVIIITGGGAARVAEALPPAFLAENTVRVA--- 570

RESULT 14
Q9WZY5

```

```

QY 238 DDNLIFLGM 246
Db 571 -DNLVIYGL 578

RESULT 13
Q9JWI7
ID AC Q9JWI7 PRELIMINARY; PRT; 592 AA.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE BirA bifunctional protein (EC 6.3.4.15).
GN BIRA OR NWA0357.
OS Neisseria meningitidis (serogroup A).
SC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL; ALJ162752; CAB83659.1; -.
DR HSSP; P06709; 1BIA.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BirA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03039; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00121; BirA_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW Ligase; Complete proteome.
SQ SEQUENCE 592 AA; 64468 MW; 7ED2E0CD3B31C630 CRC64;

Query Match 10.3%; Score 141; DB 16; Length 592;
Best Local Similarity 27.7%; Pred. No. 0.0012;
Matches 69; Conservative 38; Mismatches 98; Indels 44; Gaps 13;

QY 26 GNSRLHWAYCS-----GNAPLQWTWTDYNPKSAQLPVL-GKVPMLASVYPE--QTE 75
Db 346 GNSRLKWAVENTGTATVGSAPYR-----DLSPLGAWEAKADGNVRVGCVCGEFKAQ 401

QY 76 VWRVYQPKILTAKN-----LPLVNLV-----PSFGIDRALAGLGTGLTYGFCPLVVDGGTALT 128
Db 402 VOELARKIEWLPSSAQAALGIRNHYRHPPEHSGSDRWENALGSRFRSENACVVVSCGTAVT 461

QY 129 ITGFDQDKLVGGAIPGLGLOLATLGDRLAALPKLEMDQLTLPDRWALDTP--SAIFS 186
Db 462 VDALTDDGHVLTGTPGPFHLMKESLAVRTANLNR-----HAGKRYPPPTTTGNVAVS 514

QY 187 GVVYGVGLGALQSLQDWQKLPFGAA-----WVITGGD-GKILH-----GFLKEHSPNLSVAV 237
Db 515 GMDAVCGSVN--MMHGRLEKTKGAGKVPDVIIITGGGAARVAEALPPAFLAENTVRVA--- 570

RESULT 14
Q9WZY5

```

ID Q9W2Y5 PRELIMINARY; PRT; 246 AA.
AC Q9W2Y5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
GN TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
RL EMBL: AE001754; RAD35964.1;
DR TIGR: TM0883;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 10.1%; Score 138; DB 16; Length 246;
Best Local Similarity 25.2%; Pred. No. 0.00068;
Matches 55; Conservative 28; Mismatches 75; Indels 60; Gaps 6;
QY 66 LASVPEQTEV-----WRVYQKILTLKLNPLVNLPSFGI 101
DB 59 VASVYPTQNTVIERFSQRYFHISPIWVAKKNCVKWYKNPS-----EVGA 104
QY 102 DRALAGLTGLTYGFCPLVDGGTALTITGPDOKKLV-----GGAILPGIGLQLATIG 155
DB 105 DRVANVAVFREYKNGIIIDMGTTTV-----DLVNGSYEGGAILPGFFMWHSLF 157
QY 156 DLRAALPKLEMDQLTELPDRWALDTPSAIFSGVYGLGALQSYLQDWKLPFGAAMVIT 215
DB 158 RGTAKLPLVEKPADFVVGK---DTEENIRLGWNGSVYALRGIIIRIKVEYGDLPVILT 214
QY 216 GGDGKILHGLFLEKHSFNLVSVWDDNLIPLGMAAIIHGD 253
DB 215 GGQSKIVKDMIKHE-----IFDEDLTIKGVYHFCFGD 246

RESULT 15
Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001507; BAB03805.1;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;
Query Match 10.1%; Score 137.5; DB 16; Length 254;
Best Local Similarity 26.9%; Pred. No. 0.00078;
Matches 47; Conservative 35; Mismatches 70; Indels 23; Gaps 6;
QY 64 LMLASVPEQ-----TEVWRVY-----QPKILTLKLNPLVNLPSFGIDRALAGLGTG 111
DB 60 IVISSVVPMMFSLQDMCKKYPHVTPMIIGPGIKTGLNLIKYN-PKEVGADRIVNAVAI 118
QY 112 LTYGPFCLVVDGGTALTITGPDOKKLVGGAILPGIGLQLATIGDLRAALPKLEMDQLTE 171
DB 119 ELYGYPALVVDGFTATTTCYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIETAKPKQ 178
QY 172 LPDRWALDP-PSAIFSGVY---GVLGALQSYLQDWKLPFGAAMVITGGDGKIL 222
DB 179 VVGTTIDSMQSGIFGYVSQVDGVVYKMKQAQSEPK-----VIATGGLAKLI 227

Search completed: June 24, 2003, 21:59:30
Job time : 22.961 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.50777 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-13
Perfect score: 1368
Sequence: 1 METSKPGCLALNDKQKPW.....DDNLIFLGMAAIIHGGDRPIC 257

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.5	5.8	501	US-09-367-206-3	Sequence 3, Appli
2	76	5.6	6095	US-09-144-085-2	Sequence 2, Appli
3	75	5.5	412	US-09-134-001C-3949	Sequence 3949, Ap
4	75	5.5	561	US-08-358-901-2	Sequence 2, Appli
5	75	5.5	561	US-08-566-347-2	Sequence 2, Appli
6	75	5.5	561	US-08-693-835-2	Sequence 2, Appli
7	74.5	5.4	528	US-08-403-852D-21	Sequence 21, Appli
8	74.5	5.4	528	US-08-510-646B-22	Sequence 22, Appli
9	74.5	5.4	528	US-09-231-818-21	Sequence 21, Appli
10	74.5	5.4	4472	US-08-804-227C-2	Sequence 2, Appli
11	74	5.4	1026	US-08-194-290-7	Sequence 7, Appli
12	74	5.4	1026	US-08-614-377A-7	Sequence 7, Appli
13	74	5.4	1026	US-09-142-648B-7	Sequence 7, Appli
14	73.5	5.4	866	US-08-386-727-8	Sequence 8, Appli
15	73.5	5.4	866	US-08-600-452A-8	Sequence 8, Appli
16	73	5.3	393	US-09-432-470-2	Sequence 2, Appli
17	73	5.3	393	US-09-432-470-4	Sequence 4, Appli
18	73	5.3	655	US-09-134-001C-5531	Sequence 5531, Ap
19	72.5	5.3	474	US-09-332-041-5	Sequence 5, Appli
20	72.5	5.3	491	US-09-362-899-3	Sequence 3, Appli
21	72.5	5.3	932	US-09-071-035-416	Sequence 416, App
22	72.5	5.3	969	US-09-071-035-414	Sequence 414, App
23	72	5.3	498	US-09-172-952-26	Sequence 26, Appli
24	72	5.3	1078	US-08-480-474-11	Sequence 11, Appli
25	72	5.3	2050	US-08-347-594A-2	Sequence 2, Appli
26	71.5	5.2	239	US-08-860-174A-2	Sequence 2, Appli
27	71.5	5.2	575	US-07-683-957B-2	Sequence 2, Appli

28	71.5	5.2	943	2	US-08-808-982-7	Sequence 7, Appli
29	71.5	5.2	943	4	US-09-306-902A-7	Sequence 7, Appli
30	71	5.2	638	1	US-08-295-814E-13	Sequence 13, Appli
31	71	5.2	638	1	US-08-240-783B-2	Sequence 2, Appli
32	71	5.2	638	3	US-09-084-813-2	Sequence 13, Appli
33	71	5.2	638	4	US-09-343-361-13	Sequence 2, Appli
34	71	5.2	638	5	PCT-US92-09662-2	Sequence 2, Appli
35	71	5.2	662	4	US-09-134-001C-4074	Sequence 4074, Ap
36	71	5.2	933	2	US-08-313-200-1	Sequence 1, Appli
37	71	5.2	933	5	PCT-US93-03837-1	Sequence 1, Appli
38	70.5	5.2	249	4	US-09-372-422A-30	Sequence 30, Appli
39	70.5	5.2	297	2	US-09-027-013-3	Sequence 3, Appli
40	70.5	5.2	297	3	US-09-244-233-3	Sequence 3, Appli
41	70.5	5.2	879	1	US-08-306-546C-2	Sequence 2, Appli
42	70.5	5.2	879	2	US-08-530-524A-2	Sequence 2, Appli
43	70.5	5.2	1110	1	US-08-118-441-29	Sequence 29, Appli
44	70.5	5.2	1110	3	US-08-338-579A-29	Sequence 29, Appli
45	70.5	5.2	1110	5	PCT-US94-09851-29	Sequence 29, Appli

ALIGNMENTS

RESULT 1
US-09-367-206-3
; Sequence 3, Application US/09367206
; Patent No. 6326482
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: NSP Molecules
; FILE REFERENCE: P1223R1E
; CURRENT APPLICATION NUMBER: US/09/367,206
; CURRENT FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/08847
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: US 60/082,767
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 35
; SEQ ID NO 3
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-367-206-3

Query Match	5.8%	Score 79.5;	DB 4;	Length 501;
Best Local Similarity	19.0%	Pred. NO. 4.6;		
Matches	36;	Conservative 31;	Mismatches 95;	Indels 27; Gaps 4;
QY	13	DNDKOKPWGLMIGNSRLHWAYCSGNAPLOTWVTYDNPKSQOLPVLGKVPPLMLASVYPE	72	
Db	156	DDDRPWPPEAAQOMKQWQDKGERVTPLETTVSSFRNFEKSKFLPENKPLETAMUKR	215	
QY	73	QTEVRYVQPKILTKNLPL-VNLYPSFGIDRALAGLTGTGYPGCPVLVDGGTALTITG	131	
Db	216	AKELFTNDPDKVIAQHVLSDMCRVARIILGVSEEMR-RNMGVSSGLELITLPHGHQLRLDI	274	
QY	132	FDQKKLVGGAILPGLGLQLATIGDRILAALPK-----LEMD	167	
Db	275	IERNTMAIGAVDILGC-TGTLEDRAATLSKIIQVAVELKDSMGDLSYFSALMKALEMP	333	
QY	168	QLTELPODFW	176	
Db	334	QITRLKTKW	342	

RESULT 2
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes

APPLICANT: Betlach, Mary C.
APPLICANT: Ashley, Gary
APPLICANT: Julien, Bryan
APPLICANT: Ziermann, Rainer
TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
TITLE OF INVENTION: THEREFOR
FILE REFERENCE: 30062-20020.20
CURRENT APPLICATION NUMBER: US/09/144,085
PRIOR FILING DATE: 1998-08-31
EARLIER APPLICATION NUMBER: 09/010,809
PRIOR FILING DATE: 1998-01-22
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 6095
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-144-085-2

Query Match 5.6%; Score 76; DB 4; Length 6095;
Best Local Similarity 22.4%; Pred. No. 4.4e+02;
Matches 58; Conservative 27; Mismatches 94; Indels 80; Gaps 10;

QY 47 DYNPKSAQLPVLGKVLMLASVVPQTEVWRVYQPKILTKNLPLV-----93
DB 5856 DYASHSAHMDALLPELGAKLSLRPKATQ-----LPFYSTVTGEVSRGEAL 5901
QY 94 -----NLVPSFGIDRALAGL---GTGL---TYGPECL-----VVDGGTAL 127
DB 5902 DGEWCNRLQTVRLDRALSLELLEDHGCVFVSAHPVLAAPLTACGEAQGVVGSQR 5961
QY 128 TTGFDQDKLVGGAILPGLGLQATL---GDLAALPKLEMDQLTPELDRWALDTPSA 183
DB 5962 DEGLSLQYRTIGLHVQGHVDWARVLSGHGRAVELPYAFQR-----QRYWLDISKA 6016
QY 184 IFSGVYGVGALQSYLDQKQLPFGAAMVITGDKIL-----HGFLKHS---PNL 233
DB 6017 RSDVSSAGLKAHAAPLL-----GAATKLAEGDGHFTGRSLSGHAWLRDHEVFGNL 6068
QY 234 SVAWDDNLIFLGMAAIIHG 252
DB 6069 VFPPARGMLELALAAHPGH 6087

RESULT 3
US-09-134-001C-3949
Sequence 3949, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GPC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3949
LENGTH: 412
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3949

Query Match 5.5%; Score 75; DB 4; Length 412;
Best Local Similarity 17.6%; Pred. No. 11;
Matches 51; Conservative 45; Mismatches 87; Indels 106; Gaps 15;

QY 20 W-LGLMIGN-----SRLL-----WAYCSGNAPLOTWV-----TDYN--PKSAQLP-- 56
DB 159 WLLGLLVGWIMLLIKFHETRFALMALVVLIAWLYVFNINLTNPKVKAQLKQI 218

QY 57 -----VLIGKVPMLASVVPQTEVWRVYQPKILTKNLPLVNLVPSFGID 102
DB 219 VDVORHLILPGLGLOGAIAALVPLPK-----YATOVVKVSTV-----GYTVA 264
QY 103 RALAGLTGTLYGFPCLVVDGGTALTITGFDQDKLVGGAILPGLGLQATLGLDRLAALP 162
DB 265 IILGGIGCAFSMLFLSKIID-----NNSKGFMYGVIFSGFILTILI----- 306
QY 163 KLEMDQLTPELDRWALDTPSAIFSGVYGVGALQSYLDQKQLPFG----- 209
DB 307 -FGLSTITNIYIYVWAI-----GLFGLMYGIL-----LPWNTFMAGHINPNBOETWGV 355
QY 210 -----AAMVITGGDKILHGFLEKHSNPL-SVAMDDNLIFLGMAAIIH 250
DB 356 FNSVQGFSGMI-----GPLVGGILITQFTNLTNLTFFYSAMIFLALAVFY 399

RESULT 4
US-08-358-901-2
Sequence 2, Application US/08358901
Patent No. 5554521
GENERAL INFORMATION:
APPLICANT: Suslow, Trevor V.
APPLICANT: Jones, Jonathan D.G.
TITLE OF INVENTION: No. 5554521el Chitinase-Producing Plants
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,901
FILING DATE: 19-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/550,253
FILING DATE: 09-JUL-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/888,033
FILING DATE: 18-JUL-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/593,691
FILING DATE: 26-MAR-1984
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 12176-5-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-358-901-2

Query Match 5.5%; Score 75; DB 1; Length 561;
Best Local Similarity 22.4%; Pred. No. 17;
Matches 68; Conservative 30; Mismatches 105; Indels 100; Gaps 17;

QY 7 GCGALDNDKQKFWIGLMTGNS-----RLHWAYCSGN-----APLOTWVTDYNP 50

Db 74 GTGPKILLNGKEANSGPSTGSGTANFKVKNKGRYQMOVALCNADGCTASDATEIIVVADT 133
Qy 51 KSAQPLVLLGKVPML-----LASVVPQTEVWRVYQPKILTKNLPLVNLVPSFGI 101
Db 134 DGSHPPL--KEPPEKNKPKYKQNSKGVGSYFVEWGY--GRNFTVDKIPAOQL-----184
Qy 102 DRALAGLTGTLYGFPCLVVDGGTALTITGPDQDKLVGGAILPGLGLQLATLG-----155
Db 185 -----THLYGF--IPICGG-----NGINDSLKEIEGSF--QALQSCQCGREDPKI 226
Qy 156 ---DRLAALPKLE-----MDQLTEL-----PD-----RWALDTPSAIFSG 187
Db 227 SIHDPFAALQAKQAGVTAWDDPYKGNFGQLMALKQAHPLDKILPSIGGWTLSDPFFFMGD 286
Qy 188 VVY--GVLGALQSYLDQWKLFPGAAM--VITGGDGKILHGLFKEHSESNLSVWMDNLIF 243
Db 287 KVKRDRFVGSYKEFLQW--KEFDGVDIDWEPFGKGA-----NPNLGSPODGETYV 336
Qy 244 LGM 246
Db 337 LLM 339

RESULT 5

US-08-566-347-2
; Sequence 2, Application US/08566347
; Patent No. 5633450
; GENERAL INFORMATION:
; APPLICANT: Suslow, Trevor V.
; APPLICANT: Jones, Jonathan D.G.
; TITLE OF INVENTION: No. 5633450el Chitinase-Producing Plants
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/566.347
; FILING DATE: 01-DEC-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/358,901
; FILING DATE: 19-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/550,253
; FILING DATE: 09-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/888,033
; FILING DATE: 18-JUL-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/593,691
; FILING DATE: 26-MAR-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 12176-5-3
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-566-347-2

Query Match 5.5%; Score 75; DB 1; Length 561;
Best Local Similarity 22.4%; Pred. No. 17; Mismatches 100; Gaps 17;
Matches 68; Conservative 30;
Qy 7 CGCALDNDKQKPMGLMIGNS-----RLHWAYCSGN---APLQWTWTDYNP 50
Db 74 GTGPKILLNGKEANSGPSTGSGTANFKVKNKGRYQMOVALCNADGCTASDATEIIVVADT 133
Qy 51 KSAQPLVLLGKVPML-----LASVVPQTEVWRVYQPKILTKNLPLVNLVPSFGI 101
Db 134 DGSHPPL--KEPPEKNKPKYKQNSKGVGSYFVEWGY--GRNFTVDKIPAOQL-----184
Qy 102 DRALAGLTGTLYGFPCLVVDGGTALTITGPDQDKLVGGAILPGLGLQLATLG-----155
Db 185 -----THLYGF--IPICGG-----NGINDSLKEIEGSF--QALQSCQCGREDPKI 226
Qy 156 ---DRLAALPKLE-----MDQLTEL-----PD-----RWALDTPSAIFSG 187
Db 227 SIHDPFAALQAKQAGVTAWDDPYKGNFGQLMALKQAHPLDKILPSIGGWTLSDPFFFMGD 286
Qy 188 VVY--GVLGALQSYLDQWKLFPGAAM--VITGGDGKILHGLFKEHSESNLSVWMDNLIF 243
Db 287 KVKRDRFVGSYKEFLQW--KEFDGVDIDWEPFGKGA-----NPNLGSPODGETYV 336
Qy 244 LGM 246
Db 337 LLM 339

RESULT 6

US-08-693-835-2
; Sequence 2, Application US/08693835
; Patent No. 5776448
; GENERAL INFORMATION:
; APPLICANT: Suslow, Trevor V.
; APPLICANT: Jones, Jonathan D.G.
; TITLE OF INVENTION: No. 5776448el Chitinase-Producing Plants
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/693.835
; FILING DATE: 01-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/358,901
; FILING DATE: 19-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/550,253
; FILING DATE: 09-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/888,033
; FILING DATE: 18-JUL-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/593,691
; FILING DATE: 26-MAR-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 12176-5-3
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-693-835-2

```

Query Match      5.58; Score 75; DB 1; Length 561;
Best Local Similarity 22.4%; Pred. No. 17;
Matches 68; Conservative 30; Mismatches 105; Indels 100; Gaps 17;

QY      7 GCGLALNDKQKWLGLMIGNS-----RLHWAYCSN---APLQTWTDYNP 50
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      74 GTGPKILLNCKEAWSGSTGGTANFKVKNKGRYQMVALCNADGCTASDATEIVVADT 133
      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     51 KSAQLPVLGKVPFLM-----LASWPEQTEWVRVYQPKILTAKNPLVNLVPSFGI 101
      : : : : : : : : : : : : : : : : : : : : : : : : : :
134 DGSHLPPL--KEPLLEKNKPKYKQNSGKVGSYFVEWGVY--GRNFVDKIPQNL----- 184
      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     102 DRALAGTGLTYGFPCLVVDGGTALITGFDQDKLVGGAILPGLGLQIATLG----- 155
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     185 -----THLLYGF--IPICGG-----NGINDSLKEIEGSF---QALQSCQGRDFKI 226
      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     156 ---DRLAALPKLE-----MDQLTEL----PD-----RWALDTSATFSG 187
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     227 SIHDPPFAALQAKQAGVTAWDDPKYKGNFGQMLAKQAHPDLKILPSIGGWTLSDPFFFMGD 286
      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     188 VVY--GVIGALQSYLVQDWOKLFFPGAA--VITGGDGKILHFLKEHPNLSVAMDDNLIF 243
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     287 KVRDRFVGSVKFELQTV-KFFDGVQIDWEEFPGKGK-----NPNLGSPODGETYV 336
      : : : : : : : : : : : : : : : : : : : : : : : : : :

244 LGM 246
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     337 LLM 339
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7
US-08-403-852D-21
; Sequence 21, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,852D
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR 93/009323
 FILING DATE: 25-SEP-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 92/11441
 FILING DATE: 25-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Meyers, Kenneth J.
 REGISTRATION NUMBER: 25,146
 REFERENCE/DOCKET NUMBER: 03806.0054
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 528 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-043-852D-21

Query Match	5.4%	Score 74.5;	DB 2;	Length 528;
Best Local Similarity	21.6%;	Pred. No. 17;		
Matches	54;	Conservative	39;	Mismatches 90; Indels 67; Gaps 11;
QY	11	ALDNDRKQKW-----LGLMIGNSRLLHWAYCGSNAPLQTTWTDYDNPKSAQLPVLLG	60	
Db	27	AADRDPRR-WVILGVICLAQLVLLDNTVL-----NVAIPVLTDLGASTADIQWMIN	78	
QY	61	KVPLMLASVVPGETEWRVYQPKLILKLNPLV-----NLYPSFGID-----RALAGLGTGL	112	
Db	79	AYALVOSGLLLTAGSLADRYGKRLLMLGLVLFPGAGSAWAAFAQDSLAQLTAARAGMGVG-	137	
QY	113	TYGFPCILVVDGFTALITGFDQDKK-----LVGGAILPG-----	146	
Db	138	-----GALLATTTLAVIMQVDFDDERPRAIGLWGAASSLFGAAGPLLLGALLDHFWMGSIF	193	
QY	147	-LGLQLATIGDRLAALPKLEMDQITELPDRAWLDTSAIFS-----GVYGVVLCALQSYL	200	
Db	194	LINLPVALLG--LLAVARLVE--TRNPEGRRPDLIGAVLSTLGMVGVVYAIISGPEHGW	249	
QY	201	QDWOKLFPFGA	210	
Db	250	TAPOVLLPAA	259	

```

RESULT 8
US-08-510-646B-22
: Sequence 22, Application US/08510646B
: Patent No. 6077699
:
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Blanchet, Francis
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: APPLICANT: De Crecy-Lagard, Valer
:
: TITLE OF INVENTION: Polypeptides
: TITLE OF INVENTION: Biosynthesis
: TITLE OF INVENTION: Coding For Th
:
: NUMBER OF SEQUENCES: 45
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson,
: STREET: 1300 I Street, N.W., Su
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510.646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403.852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-510-646B-22

Query Match 5.4%; Score 74.5; DB 3; Length 528;
Best Local Similarity 21.6%; Pred. No. 17;
Matches 54; Conservative 39; Mismatches 90; Indels 67; Gaps 11;
QY 11 ALDNDKQKPW-----LGLMIGNSRLHWAYCSGNAPLOTWTDYNPKSAQLPVLLG 60
DB 27 AADRPRR-WVLGVICLAQLVLLDNTVL-----NVAIPVLTDLGASTADIQWMIN 78
QY 61 KVPMLASVPEQTEVWRVYQPKILTILKNPLV---NLYPSFGID-----RALAGLGTGL 112
DB 79 AYALVQSGLLTAGSLADRYGRKRLMLGLVLFAGSAAWAFQAQSLIARAGMGVG- 137
QY 113 TYGFPCLVVDGGTALTITGFDQDKK-----LVGGAILPG----- 146
DB 138 ----GALLATTTLAVIMQVFDDEPRRAIGLWGAASSLGFAGPLLGALLDHFHWGSIF 193
QY 147 -LGLQATLGDRLAALPKLEMDQLTELPDRWALDTPSAIFS-----GVYGVLGALQSYL 200
DB 194 LINLPVALLG--LLAVARLVE--TKNPEGRRPDLGAVLSTLGMVGVVYAIISGPEHW 249
QY 201 QDWQKLPFGA 210
DB 250 TAPQVLLPAA 259

RESULT 9
US-09-231-818-21
Sequence 21, Application US/09231818
Patent No. 6171846
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231.818
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403.852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-231-818-21

Query Match 5.4%; Score 74.5; DB 4; Length 528;
Best Local Similarity 21.6%; Pred. No. 17;
Matches 54; Conservative 39; Mismatches 90; Indels 67; Gaps 11;
QY 11 ALDNDKQKPW-----LGLMIGNSRLHWAYCSGNAPLOTWTDYNPKSAQLPVLLG 60
DB 27 AADRPRR-WVLGVICLAQLVLLDNTVL-----NVAIPVLTDLGASTADIQWMIN 78
QY 61 KVPMLASVPEQTEVWRVYQPKILTILKNPLV---NLYPSFGID-----RALAGLGTGL 112
DB 79 AYALVQSGLLTAGSLADRYGRKRLMLGLVLFAGSAAWAFQAQSLIARAGMGVG- 137
QY 113 TYGFPCLVVDGGTALTITGFDQDKK-----LVGGAILPG----- 146
DB 138 ----GALLATTTLAVIMQVFDDEPRRAIGLWGAASSLGFAGPLLGALLDHFHWGSIF 193
QY 147 -LGLQATLGDRLAALPKLEMDQLTELPDRWALDTPSAIFS-----GVYGVLGALQSYL 200
DB 194 LINLPVALLG--LLAVARLVE--TKNPEGRRPDLGAVLSTLGMVGVVYAIISGPEHW 249
QY 201 QDWQKLPFGA 210
DB 250 TAPQVLLPAA 259

RESULT 10
US-08-804-227C-2
Sequence 2, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

Search completed: June 24, 2003, 21:49:28
Job time : 9.50777 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.7173 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-13
Perfect score: 1368
Sequence: 1 METSKPGGLALDNDKQKPW.....DNLIFGLMAIHHGRPIC 257

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1368	100.0	257	9	US-09-813-453A-13
2	174	12.7	241	9	US-09-813-453A-63
3	174	12.7	265	9	US-09-813-453A-4
4	163	11.9	258	9	US-09-813-453A-2
5	158	11.5	233	9	US-09-813-453A-17
6	156.5	11.4	262	9	US-09-813-453A-45
7	152	11.1	244	9	US-09-813-453A-41
8	150	11.0	258	9	US-09-813-453A-49
9	147	10.7	260	9	US-09-813-453A-51
10	146.5	10.6	257	9	US-09-813-453A-53
11	144.5	10.6	256	9	US-09-813-453A-55
12	144	10.5	219	9	US-09-813-453A-57
13	141.5	10.3	223	9	US-09-895-913A-74
14	141.5	10.3	223	9	US-09-813-453A-14
15	141.5	10.3	223	9	US-09-813-453A-67
16	141	10.3	592	9	US-09-813-453A-22
17	141	10.3	592	9	US-09-813-453A-43
18	138	10.1	246	9	US-09-813-453A-9
19	137.5	10.1	254	9	US-09-813-453A-47

20	137	10.0	460	9	US-09-813-453A-39	Sequence 39, Appl
21	136.5	10.0	273	9	US-09-813-453A-10	Sequence 10, Appl
22	133	9.7	229	9	US-09-813-453A-12	Sequence 12, Appl
23	126	9.2	255	9	US-09-813-453A-7	Sequence 7, Appl
24	123.5	9.0	262	9	US-09-813-453A-8	Sequence 8, Appl
25	121.5	8.9	250	9	US-09-813-453A-3	Sequence 3, Appl
26	118	8.6	249	9	US-09-813-453A-61	Sequence 61, Appl
27	117	8.6	258	9	US-09-813-453A-6	Sequence 6, Appl
28	116.5	8.5	209	9	US-09-813-453A-21	Sequence 21, Appl
29	112.5	8.2	249	9	US-09-813-453A-20	Sequence 20, Appl
30	107	7.8	248	9	US-09-813-453A-20	Sequence 20, Appl
31	106	7.7	272	9	US-09-712-363-276	Sequence 276, App
32	106	7.7	272	9	US-09-813-453A-5	Sequence 5, Appl
33	99	7.2	242	9	US-09-813-453A-65	Sequence 65, Appl
34	97	7.1	267	9	US-09-813-453A-15	Sequence 15, Appl
35	92.5	6.8	212	9	US-09-813-453A-59	Sequence 59, Appl
36	90	6.6	262	9	US-09-813-453A-11	Sequence 11, Appl
37	82.5	6.0	513	10	US-09-815-242-11823	Sequence 11823, A
38	82.5	6.0	894	10	US-09-741-669-434	Sequence 434, App
39	82	6.0	305	10	US-09-815-242-13291	Sequence 13291, A
40	82	6.0	305	10	US-09-815-242-13553	Sequence 13553, A
41	82	6.0	336	9	US-09-738-626-4819	Sequence 4819, Ap
42	82	6.0	607	9	US-10-222-009-6	Sequence 6, Appl
43	82	6.0	610	9	US-10-222-009-10	Sequence 10, Appl
44	82	6.0	677	10	US-09-815-242-12100	Sequence 12100, A
45	82	6.0	679	10	US-09-815-242-5658	Sequence 5658, Ap

ALIGNMENTS

RESULT 1
US-09-813-453A-13
; Sequence 13, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: CGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Synecocystis sp.
US-09-813-453A-13
Query Match 100.0%; Score 1368; DB 9; Length 257;
Best Local Similarity 100.0%; Pred. No. 8e-121;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METSKPGGLALDNDKQKPWLGIMTGNRLHWAYCSGNAPLOTWTYDYNPKSAQLPVLIG 60
|||||
Db 1 METSKPGGLALDNDKQKPWLGIMTGNRLHWAYCSGNAPLOTWTYDYNPKSAQLPVLIG 60
QY 61 KYPLMLASVVPQTEVWRVYQPKITLKNLPLNLYPSGIDRALAGLTGLTYGPPCLV 120
Db 61 KYPLMLASVVPQTEVWRVYQPKITLKNLPLNLYPSGIDRALAGLTGLTYGPPCLV 120
QY 121 VDGGTALITGTDQDKLVGGAILPGLGLQATLGLDRLAALPKLEMDQLTLPDRWALDT 180
Db 121 VDGGTALITGTDQDKLVGGAILPGLGLQATLGLDRLAALPKLEMDQLTLPDRWALDT 180
QY 181 PSAIFSGVYVYGLGALQSYLDQWOKLFFCAAMVITGSGDKLHGFKEHSPNLSYAWDDN 240
|||||

```

RESULT 3
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Vocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 265
; TYPE: PRT

```

Query Match	11.9%	Score 163;	DB 9;	Length 258;
Best Local Similarity	28.0%;	Pred. No. 1.7e-07;		
Matches	59;	Conservative 42;	Mismatches 79;	Indels 28; Gaps 10;
Qy	64	LMLASVPE-----QTEWRYOYOPKIL-----TLKNLPLVNLVPSFGIDRALACLG	110	
Db	60	IISSVPPPIFMALERMCTRIYFI-EQIVGPMKTKGLNIKYDN-PKEVGADRIIVNAAV	117	
Qy	111	GLTYGFPCLVVDGGTATITFGDDKVLGVGATLPGLGQLATLGDRLAALPKLEMDQLT	170	
Db	118	IHLVGNPLIVVDGTTATYCYIDENQYMGGAATPGTISTEALYSRAAKLPRIETIR--	175	
Qy	171	ELPDR-WALDTPSAISGVVYGVGLGAQSLQYLQ--DRQKLPFGAAWITGCGKTLHGFLK	227	
Db	176	--PDNIIGRTVSAMQSGILFGVQVEGIVKRMKQAK--ODLKVIATGQ-----LAPLIA	228	
Qy	228	EHPNLSVAWDDNLIIFLGMAAIIHGGD	254	
Db	229	NESDCDIV-DPFLTKLGLLIYERNR	254	

RESULT 5

US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRP
; ORGANISM: Bacillus subtilis
US-09-813-453A-17

Query Match 11.5%; Score 158; DB 9; Length 233;
Best Local Similarity 30.6%; Pred. No. 4.4e-07;
Matches 48; Conservative 34; Mismatches 53; Indels 22; Gaps 7;

QY 64 LMLASVPE-----QTEWRYVQPKIL-----TKNLPVNLVPSFGIDRALAGLGT 110
DB 60 IHSVVPPIWFALECMCTKYFHI-EPOIVGPMKTLNLIKYN-PKEVGADRIVNAVA 117
QY 111 GLTYGFCPLVVDGGTALTITGFDQDQKLVGAILPGLGLQATLGDRLAALPKLEMDQLT 170
DB 118 IHLGNPLIVDFGTATTYCYIDENKQYMGAIAPGITISTEALYSRAKLPRIETR-- 175
QY 171 ELPDR-WALDTPSAIFSGVYVGLGALQSYLQ--DWQ 204
DB 176 --PDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKQ 210

RESULT 6

US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRP
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match 11.4%; Score 156.5; DB 9; Length 262;
Best Local Similarity 31.8%; Pred. No. 7.1e-07;
Matches 48; Conservative 32; Mismatches 54; Indels 17; Gaps 7;

QY 79 VYQPKILTKNLPVNLVPSFGIDR---ALAGLTGTYGFCPLVVDGGTALTITGFDQ 134
DB 87 VVPGIKTGLNIKEN--PREVGADRIVNAVAGIH---LYGSPLIIVDFGTATTYCYINE 141

RESULT 7

US-09-813-453A-41
; Sequence 41, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 244
; TYPE: PRP
; ORGANISM: Porphyromonas gingivalis
US-09-813-453A-41

Query Match 11.1%; Score 152; DB 9; Length 244;
Best Local Similarity 27.5%; Pred. No. 1.7e-06;
Matches 42; Conservative 28; Mismatches 77; Indels 6; Gaps 2;

QY 98 SFGIDRALAGLTGTYGFCPLVVDGGTALTITGFDQDQKLVGAILPGLGLQATLGD 156
DB 92 TLGADRLAALVVAHSLYPTNTELLVIDAGTATYERVSAGIYLGNIISFGLHLRKAHL 151
QY 157 LLAALPKLEMDQLTLPDRWALDTPSAIFSGVYVGLGALQSYLQDWKLPFGAAWITG 216
DB 152 FTGRLPLIDPSGIPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEGRSVILT 211
QY 217 GDGKILHGFLEKHSNLSVANDNLIIFLGMAI 249
DB 212 GDANYLARIIRS-----GILHPDLVLLGLNRI 239

RESULT 8

US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRP
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

123 APSILVDRGTATTTCATSEKGEYLGTTIAPGKISSEALFQASKLPRVE---LAKPGMT 179
176 WALDTPSAIFSGVYVGLGALQSYLDQWOKLF--PGAAMVITGGDKILHGLKHSFNL 233
180 ICKSVSAMQSGIYGVGLVDKLIISINKKELNCDVVKVIATGGLAKLI-----ASETK 233
234 SVAWDDNLIFL-GMAAIIH 250
234 SIDYVDGFLTLEGLRIY 251

RESULT 12
US-09-813-453A-57
; Sequence 57, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Dehalococcoides ethenogenes
US-09-813-453A-57

Query Match 10.5%; Score 144; DB 9; Length 219;
Best Local Similarity 23.4%; Pred. No. 8.5e-06;
Matches 50; Conservative 37; Mismatches 75; Indels 52; Gaps 7;
QY 25 IGNSRLHWAYCGSNAPLOTW-----VTDYNPKSAQLPVLGKVP-----L 64
DB 10 IGNTSVNIGITEGKLLANHLGWSVAORMADEY-----ASLLGLQHAGIHPEELNRV 63
QY 65 MLASVVPBQT---EVMRVYQPKILTAKNPLV-----NLYPSFGIDRALAG 107
DB 64 IMCVSVPLTTTTEEFKSY-----FKAAPLVVGAGIKSGVKVRMDNPREGADRIYNA 117
QY 108 LGTGLTYGFCPLVVDGDTALTITGFDQDKKLVGGAIPGLGLQLATLGDRLAALPKLEMD 167
DB 118 AAARLVYPGACIIVDMGTATFTDLSEGGVYIGGAIAPGIATSAQAIAEKTSKLPKIBII 177
QY 168 QLTLPDRWALDTPSAIFSGVYVGLGALQSYLQ 201
DB 178 RPAKV---IGNTVSMQSGIYGVGLVEELVR 208

RESULT 13
US-09-895-913A-74
; Sequence 74, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Omen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227

; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-74
Query Match 10.3%; Score 141.5; DB 9; Length 223;
Best Local Similarity 26.9%; Pred. No. 1.5e-05;
Matches 66; Conservative 28; Mismatches 76; Indels 75; Gaps 12;
QY 25 IGNSRLHWA-----YCSGNAPLOTWVTYDYNPKSAQLPVLGKVPMLMASVVPQEOTVWRV 79
DB 18 IGNTRIHFAQNYQLFSSAKEDLKR-----LGIOKEIFISVNEENE----- 58
QY 80 YQPKILTAKNPLVNLNLYPS-----FGIDRALAGLGTGLTYGFCPLVVDG 123
DB 59 -----KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVYDA 103
QY 124 GTALTITGFDQDKKLVGGAIPGLG--LQATLGDRLAALPKLEMDQLTLPDRWALDTP 181
DB 104 GSAITIDLIKEGKHL-GGCILPGLAQYTHAYKKSAKILEQPFKALDSLEVLPK-----STR 158
QY 182 SAIFSGVYVGLGALQSYLDQWOKLFGCAAMVITGGDKILHGLKHSFNLVSAWDDNL 241
DB 159 DAVNYGWLVSIVACIQ-HLAKNQIY-----LCGGDAKYLSAFL-PHS-----VCKERL 205
QY 242 IFGLM 246
DB 206 VFDGM 210

RESULT 14
US-09-813-453A-14
; Sequence 14, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-813-453A-14

Query Match 10.3%; Score 141.5; DB 9; Length 223;
Best Local Similarity 26.9%; Pred. No. 1.5e-05;
Matches 66; Conservative 28; Mismatches 76; Indels 75; Gaps 12;
QY 25 IGNSRLHWA-----YCSGNAPLOTWVTYDYNPKSAQLPVLGKVPMLMASVVPQEOTVWRV 79
DB 18 IGNTRIHFAQNYQLFSSAKEDLKR-----LGIOKEIFISVNEENE----- 58
QY 80 YQPKILTAKNPLVNLNLYPS-----FGIDRALAGLGTGLTYGFCPLVVDG 123
DB 59 -----KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVYDA 103
QY 124 GTALTITGFDQDKKLVGGAIPGLG--LQATLGDRLAALPKLEMDQLTLPDRWALDTP 181
DB 104 GSAITIDLIKEGKHL-GGCILPGLAQYTHAYKKSAKILEQPFKALDSLEVLPK-----STR 158

RESULT 15

Query Match 10.3%; Score 141.5; DB 9; Length 223;

Best Local Similarity	26.9%	Pred. No. 1.5e-05;
Matches	66;	Conservative 28; Mismatches 76;
Indels	75;	Gaps 12;

Search completed: June 24, 2003, 22:24:01
Job time : 12.7173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.4883 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-6
Perfect score: 1367
Sequence: 1 MLLCIDGNTNTFVSVDGT.....DLTMHGLRLIFDYNKGLGA 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1367	100.0	258	23 AAU91153	Rhodobacter capsul
2	658	48.1	260	23 AAU91173	Pantothenate kinas
3	554	40.5	255	23 AAU91154	Geobacter sulfurre
4	542	39.6	258	23 AAU91172	Pantothenate kinas
5	537	39.3	258	22 AAU01243	B. subtilis novel
6	537	39.3	258	23 AAU91149	Bacillus subtilis
7	536	39.2	262	23 AAU91170	Pantothenate kinas
8	528	38.6	254	23 AAU91171	Pantothenate kinas
9	527	38.6	265	23 AAU91151	Streptomyces coeli
10	496.5	36.3	256	23 AAU91175	Pantothenate kinas

11	480	35.1	259	23 ABB47661	Listeria monocytog
12	455	33.3	233	23 AAU91163	Pantothenate kinas
13	431.5	31.6	272	23 AAU91152	Mycobacterium tube
14	428.5	31.3	250	23 AAU91150	Clostridium acetob
15	427.5	31.3	272	22 AAG81225	Mycobacterium tube
16	394	28.8	219	23 AAU91176	Pantothenate kinas
17	347.5	25.4	262	23 AAU91155	Deinococcus radiop
18	295.5	21.6	246	23 AAU91156	Thermotoga maritim
19	293.5	21.5	273	23 AAU91157	Treponema pallidum
20	265	19.4	257	23 AAU91174	Pantothenate kinas
21	240.5	17.6	212	23 AAU91177	Pantothenate kinas
22	224	16.4	244	23 AAU91168	Pantothenate kinas
23	197.5	14.4	262	23 AAU91158	Borrelia burgdorfe
24	168.5	12.3	241	23 AAU91179	Pantothenate kinas
25	158.5	11.6	229	23 AAU91159	Pantothenate kinas
26	140	10.2	455	21 AAY74910	Aquifex aeolicus p
27	140	10.2	592	20 AAY38616	Neisseria meningit
28	140	10.2	592	21 AAY74913	Neisseria meningit
29	140	10.2	592	23 AAU91166	Pantothenate kinas
30	139	10.2	455	20 AAY38617	Neisseria gonorrh
31	139	10.2	455	21 AAY74908	Neisseria gonorrh
32	139	10.2	460	23 AAU91167	Pantothenate kinas
33	139	10.2	592	20 AAY38618	Neisseria gonorrh
34	139	10.2	592	21 AAY74911	Neisseria gonorrh
35	136	9.9	248	23 AAU91164	Pantothenate kinas
36	134	9.8	267	23 AAU91162	Bordetella pertussis
37	134	9.8	389	21 AAY74909	Neisseria meningit
38	134	9.8	592	20 AAY38615	Neisseria meningit
39	134	9.8	592	21 AAY74912	Neisseria meningit
40	134	9.8	592	23 AAU91169	Pantothenate kinas
41	132	9.7	249	23 AAU91182	Pantothenate kinas
42	125.5	9.2	242	23 AAU91180	Pantothenate kinas
43	124	9.1	249	23 AAU91178	Pantothenate kinas
44	117	8.6	257	23 AAU91160	Synechocystis pant
45	96	7.0	209	23 AAU91165	Pantothenate kinas

ALIGNMENTS

RESULT 1
AAU91153
ID AAU91153 standard; Protein; 258 AA.
XX
AC AAU91153;
XX
DT 05-JUN-2002 (first entry)
XX
Rhodobacter capsulatus pantothenate kinase Coax.
DE
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
KW
XX Rhodobacter capsulatus.
OS
XX WO200216601-A2.
PN
PD 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
PF
XX 24-AUG-2000; 2000US-227860P.
PR
PR 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PA Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -

XX
PS Claim 10; Page 71-72; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 258 AA;

Query Match 100.0%; Score 1367; DB 23; Length 258;
Best Local Similarity 100.0%; Pred. No. 5.7e-138;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MLLCIDCGNTNTVFSWGDGTFDAATWRIATDHRRTADEYFVWLNTLMQLKLGGRISAI 60
|||||
1 MLLCIDCGNTNTVFSWGDGTFDAATWRIATDHRRTADEYFVWLNTLMQLKLGGRISAI 60
|||||
61 ISSTAPRVVFNLRVLCNRYFDCRPYVVGKPGCELPVAPRDPGTTVGPDRLVNTVAGYDR 120
|||||
61 ISSTAPRVVFNLRVLCNRYFDCRPYVVGKPGCELPVAPRDPGTTVGPDRLVNTVAGYDR 120
|||||
121 HGGDLIVDFGTFATTFDVVAPDGVIGVAPGVNLSLEALHMAAALPHVDVTRKPGVI 180
|||||
121 HGGDLIVDFGTFATTFDVVAPDGVIGVAPGVNLSLEALHMAAALPHVDVTRKPGVI 180
|||||
181 GNTVACIQSGVYWGIVGLVEGIVRQIRMERDRPMKVATGGLASLFDLGFDFKVEDD 240
|||||
181 GNTVACIQSGVYWGIVGLVEGIVRQIRMERDRPMKVATGGLASLFDLGFDFKVEDD 240
|||||
241 LPMHGLRLIFDYNKGLGA 258
|||||
241 LPMHGLRLIFDYNKGLGA 258
|||||

RESULT 2
AAU911173
ID AAU911173 standard; Protein; 260 AA.
AC AAU911173;
XX
XX
DT 05-JUN-2002 (first entry)

Pantothenate kinase (Coax) #11.
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Caulobacter crescentus.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
PF 20-MAR-2001; 2001US-0813453.
PR
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PA
XX
XX Yocum RR, Patterson TA;
PI WPI; 2002-269358/31.
XX
DR N-PSDB; ABK54194.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
XX Claim 10; Page 102-103; 128pp; English.
PS
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
SQ Sequence 260 AA;

Query Match 48.1%; Score 658; DB 23; Length 260;
Best Local Similarity 50.8%; Pred. No. 5.9e-62;
Matches 130; Conservative 48; Mismatches 74; Indels 4; Gaps 3;
1 MLLCIDCGNTNTVFSWGDGTFDAATWRIATDHRRTADEYFVWLNTLMQLKLGGR-ISA 59
|||||
1 MLLAIEQNTNTMTFAIHDGASWVAQWRSATESRTADEYVWLSQLLSMOGLGFRAIDAV 60
|||||
60 IISSTAPRVVFNLRVLCNRYFDCRPYVVGKPGCELPVAPRDPGTTVGPDRLVNTVAGYD 119
|||||
61 IISVVSPOSIFNLRNLSRRYFNVEPLVIGE-NAKIGIDVRIEKPSEAGADRLVNAIGAAM 119
|||||
120 RHGGDLIVDFGTFATTFDVVAPDGVIGVAPGVNLSLEALHMAAALPHVDVTRKPGQ- 178
|||||
120 VYPGLVVIDSGTATTFDVAADGAFEGGIAPGINSQALHEAAKLPRATIQPAGN 179
|||||
179 -VIGNTVACIQSGVYWGIVGLVEGIVRQIRMERDRPMKVATGGLASLFDLGFDFKDV 237
:|||||
180 RIVGTDVTSAMQSGVFWGYISLIEGLVARIKAEERGEPTVIATGGVASFEGATSIDHF 239
|||||
238 EDDLTMHGLRLIFDYN 253
:|||||
240 DSDLTIRGLLEIYRN 255
:|||||

RESULT 3
AAU911154
ID AAU911154 standard; Protein; 255 AA.
XX
XX
AC AAU911154;
XX

05-JUN-2002 (first entry)
XX
XX Geobacter sulfurreducens pantothenate kinase Coax.
XX
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
KW
KW Geobacter sulfurreducens.
XX
OS WO200216601-A2.
XX
PN 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
PF
XX 24-AUG-2000; 2000US-227860P.
PR
XX 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PA
XX Yocum RR, Patterson TA;
PI WPI; 2002-269358/31.
XX
XX
DR WPI; 2002-269358/31.
XX

XX Yocum RR, Patterson TA, Hermann T, Pero JG;
 PI WPI: 2001-218644/22.
 DR N-PSDB; AAS00984.
 XX New recombinant microorganism which overexpress a *Bacillus subtilis*
 PT pantothenate biosynthetic enzyme, useful for the high yield production
 PT of panto-compounds such as pantothenate and pantoate -
 XX Example 14; Fig 23; 292pp; English.
 PS The sequence represents a novel *B. subtilis* pantothenate kinase (encoded
 XX by gene cox), an enzyme of the pantothenate biosynthetic pathway.
 CC Pantothenate, also known as vitamin B5, is used as a nutritional
 CC supplement in mammals and humans. The invention concerns methods of
 CC producing recombinant microorganisms overexpressing at least one *Bacillus*
 CC *subtilis* pantothenate biosynthetic enzyme. The microorganisms and methods
 CC of producing them are useful for producing a panto-compound such as
 CC pantothenate or pantoate, which is a nutritional requirement for
 CC livestock and humans. The methods are also useful for the identification
 CC of pantothenate kinase modulators. Panto-compounds are produced at a
 CC significantly higher yield than prior art methods and can be produced
 CC independent of the need to feed precursors which decreases expense.
 XX Sequence 258 AA;
 SQ Query Match 39.3%; Score 537; DB 22; Length 258;
 Best Local Similarity 43.9%; Pred. No. 5.5e-49;
 Matches 112; Conservative 46; Mismatches 95; Indels 2; Gaps 2;
 QY 1 MLLCICGNTNTVFSVWGDTDFAAWRIATDHRRTADEYFVWLTMLQKGLQ-GRISEA 59
 Db 1 LLLVIDVGNNTVGLVYHDKLEYHWRIETSRHKTDEFEFGMLRSLFDHSGLMFQIDGI 60
 QY 60 IISSTAPRVVFNRLVLCNRYFDCRPYVVGKPCCELPVAPVDPGTTVGDRLVNTVAGYD 119
 Db 61 IISVVPPIMFALERMCTKYFHEPQIVG-PGKMTGLNPKYDNKPEVGAADRVNVAATH 119
 QY 120 RHGGDLIVDFGTATTDFVAPDGAYIGGVIAAGVNLSEALHMAAALPHVDVTKPGCV 179
 Db 120 LYGNNLVVDFGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETRPDNI 179
 QY 180 IGTNTVACIQSGVYNGYIGLVGIVRQIRMERDRPMKVATGGLASLFDLFDKVED 239
 Db 180 IGTNTVSAMQSGILFGYGVQVEGIVKRMKQAKDLKVATGGLAPLANESDCIDIVDP 239
 QY 240 DLTMHGLRLIFDYNK 254
 Db 240 FLTLAGLELIYERNR 254
 RESULT 6
 AAU911149 standard; Protein; 258 AA.
 XX ID AAU911149
 XX AC AAU911149;
 XX DT 05-JUN-2002 (first entry)
 XX DE *Bacillus subtilis* pantothenate kinase Coax.
 XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 OS *Bacillus subtilis*.
 XX WO200216601-A2.
 XX PN 28-FEB-2002.
 XX PD 24-AUG-2001; 2001WO-US26531.
 XX PF 24-AUG-2001; 2001US-0813453.

PR 24-AUG-2000; 2000US-227860P.
 XX 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX Yocum RR, Patterson TA;
 XX WPI: 2002-269358/31.
 DR N-PSDB; ABK54168.
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 PS Claim 10; Page 67-68; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX Sequence 258 AA;
 SQ Query Match 39.3%; Score 537; DB 23; Length 258;
 Best Local Similarity 43.9%; Pred. No. 5.5e-49;
 Matches 112; Conservative 46; Mismatches 95; Indels 2; Gaps 2;
 QY 1 MLLCICGNTNTVFSVWGDTDFAAWRIATDHRRTADEYFVWLTMLQKGLQ-GRISEA 59
 Db 1 LLLVIDVGNNTVGLVYHDKLEYHWRIETSRHKTDEFEFGMLRSLFDHSGLMFQIDGI 60
 QY 60 IISSTAPRVVFNRLVLCNRYFDCRPYVVGKPCCELPVAPVDPGTTVGDRLVNTVAGYD 119
 Db 61 IISVVPPIMFALERMCTKYFHEPQIVG-PGKMTGLNPKYDNKPEVGAADRVNVAATH 119
 QY 120 RHGGDLIVDFGTATTDFVAPDGAYIGGVIAAGVNLSEALHMAAALPHVDVTKPGCV 179
 Db 120 LYGNNLVVDFGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETRPDNI 179
 QY 180 IGTNTVACIQSGVYNGYIGLVGIVRQIRMERDRPMKVATGGLASLFDLFDKVED 239
 Db 180 IGTNTVSAMQSGILFGYGVQVEGIVKRMKQAKDLKVATGGLAPLANESDCIDIVDP 239
 QY 240 DLTMHGLRLIFDYNK 254
 Db 240 FLTLAGLELIYERNR 254
 RESULT 7
 AAU911170 standard; Protein; 262 AA.
 XX ID AAU911170
 XX AC AAU911170;
 XX DT 05-JUN-2002 (first entry)
 XX DE Pantothenate kinase (Coax) #8.
 XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 OS *Bacillus anthracis*.
 XX WO200216601-A2.
 XX PN 28-FEB-2002.
 XX PD 28-FEB-2002.

PN WO200216601-A2.
 XX 28-FEB-2002.
 XX 24-AUG-2001; 2001WO-US26531.
 XX 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA Yocum RR, Patterson TA;
 PI WPI; 2002-269358/31.
 DR Identifying potential antibiotic or antimicrobial agent, comprises
 XX contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX Claim 10; Page 69-70; 128pp; English.
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX Sequence 265 AA;
 SQ Query Match 38.6%; Score 527; DB 23; Length 265;
 Best Local Similarity 45.8%; Pred. No. 6.7e-48;
 Matches 119; Conservative 41; Mismatches 93; Indels 8; Gaps 3;
 QY 1 MLICDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLTMLQKLGQ-----R 55
 DB 1 MLITIDVGNTHTVLGLFDEGDIWEHNRISTDSRRTADLAVLLQGLMGHPLLDGLDGG 60
 QY 56 ISBAIISSTAPRVFNLVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTGVDRLVNTV 115
 DB 61 IDGIALCAIVPSVHLRELVTTRYGDPVAVLVEPGVKGTPILTDHPKEVGADRIINAV 120
 QY 116 AGYDRHGGDLIVVDFGTATTEDVAPDGAIGVIAAGVNLSEALHMAAALPHVDVTK 175
 121 AAVELYGGPAIVVDFGTATTEDVAPDGAIGVIAAGVNLSEALHMAAALPHVDVTK 180
 176 PQGVIGTNTVACIQSGVYWGVIYGLVEGIVRQIRME--RDRP--MKVIATGGLASLFDL 232
 181 PRSVIGKNTVEAMQSGIYVGFAGQGVGVNRMARELADDDVTVIATGGLAPVWLGE 240
 QY 233 LFDKVEDDTMHLRLIFDYN 253
 DB 241 VIDEHEPWLTLMLGLRLVYERN 261
 RESULT 10
 AAU91175
 ID AAU91175 standard; Protein; 256 AA.
 XX AC AAU91175;
 XX DT 05-JUN-2002 (first entry)
 XX Pantothenate kinase (Coax) #13.
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX Clostridium difficile.
 OS

XX WO200216601-A2.
 XX 28-FEB-2002.
 XX 24-AUG-2001; 2001WO-US26531.
 XX 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA Yocum RR, Patterson TA;
 PI WPI; 2002-269358/31.
 DR N-PSDB; ABR54196.
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX Claim 6; Page 105; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX Sequence 256 AA;
 SQ Query Match 36.3%; Score 496.5; DB 23; Length 256;
 Best Local Similarity 41.4%; Pred. No. 1.2e-44;
 Matches 106; Conservative 49; Mismatches 98; Indels 3; Gaps 3;
 QY 1 MLICDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLTMLQKLGQ-GRISEA 59
 DB 1 MLIVDFVGNTHTVLGLFDEGDIWEHNRISTDSRRTADLAVLLQGLMGHPLLDGLDGG 60
 QY 60 IISSTAPRVFNLVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTGVDRLVNTVAGYD 119
 DB 61 IISVVPVNVHSENFCEIKYCKKQPLIVG-PGKTLGNLKYDNPKQVQADRVNAGIE 119
 QY 120 RHGGDLIVVDFGTATTEDVAPDGAIGVIAAGVNLSEALHMAAALPHVDVTKPGV 179
 DB 120 KYGAPSLVDFGTATTEDVAPDGAIGVIAAGVNLSEALHMAAALPHVDVTKPGV 179
 QY 180 IGTNTVACIQSGVYWGVIYGLVEGIVRQIRMERD-REPKVIATGGLASLFDLFDKVE 238
 DB 180 ICKSTVSAMQSGIYVGFAGQGVGVNRMARELADDDVTVIATGGLAPVWLGE 239
 QY 239 DDLTMHGLRLIFDYN 254
 DB 240 GFLEGLRLIYEKNQ 255
 RESULT 11
 ABB47661
 ID ABB47661 standard; Protein; 259 AA.
 XX AC ABB47661;
 XX DT 05-FEB-2002 (first entry)
 XX Listeria monocytogenes protein #365.
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.

AAG81225

ID AAG81225 standard; Protein; 272 AA.

XX AC AAG81225;

XX DT 04-SEP-2001 (first entry)

XX DE Mycobacterium tuberculosis potential drug target protein SEQ ID 276.

XX KW Drug target; growth; organism viability; characterisation.

XX OS Mycobacterium tuberculosis.

XX PN WO200135317-A1.

XX PD 17-MAY-2001.

XX PF 13-NOV-2000; 2000WO-US31152.

XX PR 12-NOV-1999; 99US-0165086.

XX PR 12-NOV-1999; 99US-0165124.

XX PA 01-FEB-2000; 2000US-0179531.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Eisenberg D, Rotstein SH, Marcotte EM;

XX DR WPI; 2001-329193/34.

XX DR N-PSDB; AAG81225.

XX PT Identifying nucleotide or polypeptide sequence for use as drug target, involves providing algorithm that analyzes a functional relationship between nucleotide or polypeptide sequences, and comparing the sequences

XX PS Disclosure; Page 188; 207pp; English.

XX CC This invention relates to a method for identifying a nucleotide or polypeptide sequence that may be a drug target, or essential for growth or viability of an organism. Polynucleotide sequences AAG81225 - AAG81225 represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium tuberculosis proteins which are potential drug targets. The DNA and protein sequences are used to illustrate the method of the invention. The method involves providing an unknown nucleotide or polypeptide sequence, and comparing it to a number of sequences along with at least one algorithm capable of analysing a functional relationship between nucleotide and polypeptide sequences. The method is useful for characterising the function of nucleic acids and polypeptides that may be useful as a target for a drug or essential for the growth or viability of an organism.

XX SQ Sequence 272 AA;

Query Match 31.3%; Score 427.5; DB 22; Length 272;

Best Local Similarity 37.2%; Pred. No. 3.2e-37;

Matches 99; Conservative 47; Mismatches 101; Indels 19; Gaps 4;

Qy 1 MLCIDCGNTVTFVWDG---TDFATWRIATDHRRTADEYFVWLTMLQKLGRI 56

Db 1 VLLAIDVNRHTTVVGLSGMKEHAKVQQWRIRTESEVTDE-----LALTIDGLIGED 54

Qy 57 SEALIS---STAPRVFNLRVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTVGPDRLV 112

Db 55 SERLTGTAALSTVPSVLHEVRIMLDQYWPSPVPHVLEPGVTGIPLLVDNPKVQADRV 114

Qy 113 NTVAGYDRHGDDLIVDFGATTFDVVAPDGAYIGGVVAPGVNLSLEALHMAAALPHVD 172

Db 115 NCLAAVDFRKAALVDFGSSICVDVYSAKGEFLGAIPGVQVSSDAAAARSALRRVE 174

Qy 173 VTKPQGVIGNTVACIQSGVYTWYGLVGEIVRQIR-----MERDRPKVIATGGLSLF 227

Db 175 LARPSRVVGKNTVECMQAGAVFGAGLVLDVGRVREDVSGFSVDHDAIVATGHTAPLL 234

Qy 228 DLGFDLFDKVEDDLTMHGLRLIFDYN 253
 Db 235 LPELHTVDHYDQHLTLQGLRLVFRN 260

Search completed: June 24, 2003, 21:46:14
 Job time : 27.4883 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.41033 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453a-6

Perfect score: 1367

Sequence: 1 MLLCIDCGNTVFSVMDGT.....DDLTMHGLRLIFDYNKGLGA 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	658	48.1	261	2 B87489	transcription acti
2	528	38.6	254	2 F83660	hypothetical prote
3	527	38.6	265	2 T36391	hypothetical prote
4	485	35.5	273	2 E97293	probable transcrip
5	480	35.1	259	2 AF1102	conserved hypotet
6	479	35.0	259	2 AF11464	conserved hypotet
7	455	33.3	233	2 S66100	conserved hypotet
8	431.5	31.6	272	2 A70955	hypothetical prote
9	431.5	31.6	274	2 H86937	conserved hypotet
10	347.5	25.4	262	2 E75516	conserved hypotet
11	295.5	21.6	246	2 D72320	conserved hypotet
12	293.5	21.5	273	2 D71326	conserved hypotet
13	197.5	14.4	262	2 F70165	conserved hypotet
14	158.5	11.6	229	2 E70465	hypothetical prote
15	140	10.2	592	2 H82031	probable biotin-la
16	136	9.9	248	2 H83111	hypothetical prote
17	134	9.8	267	2 I40327	baf protein - Bord
18	134	9.8	592	2 B81009	BirA protein/Bvg a
19	129	9.4	276	2 A12292	hypothetical prote
20	125.5	9.2	242	2 A82637	conserved hypotet
21	117	8.6	257	2 S75559	hypothetical prote
22	96	7.0	209	2 H81382	hypothetical prote
23	94.5	6.9	440	2 T35925	diaminopimelate de
24	92	6.7	334	2 B75095	probable histone d
25	89.5	6.5	336	2 T35643	glycerol-3-phospha
26	89.5	6.5	391	2 D97251	S-adenosylmethioni
27	89	6.5	341	1 DEETHH	L-threonine 3-dehy
28	89	6.5	341	2 F91190	threonine dehydrog
29	89	6.5	341	2 G86037	threonine dehydrog

30	85.5	6.3	225	2 A31583	C-reactive protein
31	86	6.3	411	2 A72506	probable tRNA spli
32	85.5	6.3	356	2 A95900	probable catabolit
33	85	6.2	344	2 G82690	aspartate-B-semal
34	85	6.2	398	2 F89964	S-adenosylmethioni
35	84.5	6.2	299	2 A75467	methylenetetrahydr
36	84.5	6.2	341	2 AB0974	L-threonine 3-dehy
37	84.5	6.2	385	2 A72480	hypothetical prote
38	84.5	6.2	467	2 G87564	aminopeptidase, pr
39	84	6.1	1763	2 T17465	rifamycin polyketi
40	83.5	6.1	341	2 AD0291	purine nucleotide
41	83.5	6.1	396	2 G97955	methionine adenosy
42	83.5	6.1	1747	1 A45974	collagen alpha 1(X
43	83	6.1	370	2 C90928	hypothetical prote
44	83	6.1	370	2 G85776	hypothetical prote
45	83	6.1	370	2 H64926	probable membrane

ALIGNMENTS

RESULT 1

B87489

transcription activator, probable Baf family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: B87489

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K. n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87489

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-261 <S>O>

A:Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC1935

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match	48.1%	Score 658;	DB 2;	Length 261;
Best Local Similarity	50.8%;	Pred. No. 2.8e-50;		
Matches 130;	Conservative 48;	Mismatches 74;	Indels 4;	Gaps 3;
QY	1	MLLCIDCGNTVFSVMDGTDFAATWRTATDHRRTADEYFVWLNTLMQLKGLQGR-ISEA	59	
Db	2	MLLAIEQGNNTMTFAIHGDGASWVAQWRSATSTRTADEYVWVLSQLLSMQGLGFRATDAV	61	
QY	60	IISSTAPRVFNLRVLCNRYFCRPPYVVGKPGCELPVAPRVDPGTTVPDRLVNTVAGYD	119	
Db	62	IISVVPQSIENLRNLSRYPNVEPLVIGE-NAKIGIDVRIEKPSEAGADRLVNAIGAA	120	
QY	120	RHGGLIIVDFGTATTFDVPADGAYIGGVVAPGVNLSLEALHMAAALPHVDVTKPGQ-	178	
Db	121	VYPGLVVIDSGTATTFDVAADGAFEGGIAPGINSQALHEAAAKLPRIATQRPAGN	180	
QY	179	-VIGTNTVACIQSGVWGYIGLVESIQVIRNDRPMKVIATGGLASLFDLGFDFPKV	237	
Db	181	RIVGTDVTSAMQSGVFWGISLIEGLVARIEKAERGEPTMTVIATGGVASLFEGATDSIDH	240	
QY	238	EDDLTMHGLRLIFDYN	253	
Db	241	DSLDTIRGLLEIYRRN	256	

RESULT 2

F83660

hypothetical protein BH086 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: F83660

R.Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83660
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:CROSS-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0086
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.6%; Score 528; DB 2; Length 254;
Best Local Similarity 41.3%; Pred. No. 7.1e-39;
Matches 105; Conservative 55; Mismatches 92; Indels 2; Gaps 2;
QY 1 MLLCIDCGNTNTVFSVWDGTFDAATWRIATDHRRTADEYFVWLNTLMQLGQGR-ISEA 59
1 MILVIDGNTNTVLGVYQDETLLVHHRLATSQKTEDEYAMTVRSLEFDHAGLQFDIDGI 60
60 IISSTAPRVNVLNLCNRYFDCRPYVVGKPCCELPVAPRVPDGTGTPDRLVNTVAGYD 119
61 VISSVVPMPFSLQCKKQYFHVPMITG-PGKTGLNKKYDNPKEVGADRIVNAVAIE 119
120 RRGGLIVVDFGTATTFDVPADGAYIGGVAPGVNLSLEALHMAAALPHVDVTPQGV 179
120 LXGYPAIVVDFGTATTYCLINEKKQYAGGVAPGIMISTEALYHRASKLPRIEIAKPKQV 179
180 IGTNTVACIQSGVWGYIGLVEGIVRQIRMERDRPMKVATGGLASLFDLFDKVED 239
180 VGTNTIDSMQSGIFGYVQSDVGVVVKRMAQAESEPKVATGGLAKLIGTSETIDVIDS 239
QY 240 DLTMHGLRLIFDYN 253
240 FLTLGQLIYKKN 253

RESULT 3
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36391
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T36391
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-265 <OLI>
A:CROSS-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCODB:SCE94.31c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODB:SCE94.31c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.6%; Score 527; DB 2; Length 265;
Best Local Similarity 45.6%; Pred. No. 9.1e-39;
Matches 119; Conservative 41; Mismatches 93; Indels 8; Gaps 3;
QY 1 MLLCIDCGNTNTVFSVWDGTFDAATWRIATDHRRTADEYFVWLNTLMQLGQGR-ISEA 55
1 MLLTIDVGNTHVLGLFDGEDIVHWRISTDSRRTADELAVLQLGCMGHPLLGDELGDG 60
56 ISEALISSTAPRVNVLNLCNRYFDCRPYVVGKPCCELPVAPRVPDGTGTPDRLVNTV 115
61 IDGAIATVPSVLHRELVTRRYGDPVAVLVEPGVTKGPIILTDHPKEVGADRIINAV 120
116 AGYDRHGLDIVDFTGATTFDVPADGAYIGGVAPGVNLSLEALHMAAALPHVDVTK 175
121 AAVELYGGPAIVVDFGTATTFDVPADGAYIGGVAPGVNLSLEALHMAAALPHVDVTK 180

QY 176 PQGVIGTNTVACIQSGVWGYIGLVEGIVRQIRMERDRPMKVATGGLASLFDLFD 232
181 PRSVIGKNTVEAMQSGIYGFAGQVDGVVNRARELADDDVTVIATGGLAPVWLGE 240
QY 233 LFDKVEDDLTMHGLRLIFDYN 253
241 VIDEHEPWTLMGLRLVVERN 261

RESULT 4

E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostr
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97293
R:Nolling, J.; Brenon, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273 <KUR>
A:CROSS-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3200
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 35.5%; Score 485; DB 2; Length 273;
Best Local Similarity 40.2%; Pred. No. 4.6e-35;
Matches 104; Conservative 51; Mismatches 98; Indels 6; Gaps 3;
QY 1 MLLCIDCGNTNTVFSVWDGTFDAATWRIATDHRRTADEYFVWLNTLMQLGQGR-ISEA 59
12 VILVDVGNNTNVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQDKLDPITLVEGV 71

QY 60 IISSTAPRVNVLNLCNRYFDCRPYVVGKPCCELPVAPRVPDGTGTPDRLVNTVAGYD 119
72 IISVVPNIMYSLEHMRIRKYNKINLVVG-PGKTKINIKYDNPKEVGADRIVNAVAIE 130
QY 120 RRGGLIVVDFGTATTFDVPADGAYIGGVAPGVNLSLEALHMAAALPHVDVTPQGV 179
131 IVKRSLLIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYA 190
QY 180 IGTNTVACIQSGVWGYIG-IVVEGIVRQIRMERDRPMKVATGGLASLFDLFDLFD 235
191 ICKNTISSIQSGIVGYICQVRYIVERMKEQLQEBEGEKEPLVATGGLAKLISEAKNYD 250
QY 236 KVEDDLTMHGLRLIFDYNK 254
251 VINPFLTEGURIIYKRN 269

RESULT 5

AF1102
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EG
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1102
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
Ok, C.; Schlueter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1102
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-259 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo221
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 35.1%; Score 480; DB 2; Length 259;
Best Local Similarity 38.5%; Pred. No. 1.2e-34;
Matches 101; Conservative 53; Mismatches 92; Indels 16; Gaps 4;

QY 1 MLLCIDCGNTNTVFSVMDGTDFAATWRIATDHRRTADE-----YFWLNL-TLMQLKGL 52
DQ 1 MILVIDGNTCTGVGYKEQKLLRHRMTTDRHRTSDGLMGLNLFNFSYANLTPSDIQGI 60
QY 53 QGRISEAIIISAPRVFNRLVLCNRYFDCRPYVVGKPGCELPVAPVDPGTTTVPDRVLV 112
DQ 61 -----IISVVPPIMHAMETMCVRVFNIRPLVIG-PGIKTGLNKLKVDNPREIGSDRIV 112
QY 113 NTVAGYDRHGDLIVDFGTATTEDVVPADGAYIGGVIAAGVNLNLEALHMAAALPHVD 172
DQ 113 NAVAASEEYGPVIVDEGTATTCYIDESGYOGGAIAFGIMISTEALYNRAAKLPRVD 172
QY 173 VTKPGQVIGTNVACIQSGVYWGIVGIVRQIRMERDRPMKVIATGGGLASFLDLGFD 232
DQ 173 IAESSQIIIGKSTVSSMQAGIFGYFGQCEGIIAEMKKQSNASPVVATGGLARMITERKSS 232
QY 233 LFDKVEDDLTMHGLRLIFDYNK 254
DQ 233 AVDILDPLTLKGLLELYRRNK 254

RESULT 6
AF1464
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1464
A:Authors: P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1464
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin0253
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 35.0%; Score 479; DB 2; Length 259;
Best Local Similarity 39.3%; Pred. No. 1.4e-34;
Matches 103; Conservative 50; Mismatches 93; Indels 16; Gaps 4;

QY 1 MLLCIDCGNTNTVFSVMDGTDFAATWRIATDHRRTADE-----YFWLNL-TLMQLKGL 52
DQ 1 MILVIDGNTCTGVGYKEQKLLRHRMTTDRHRTSDGLMGLNLFNFSYANLTPSDIQGI 60
QY 53 QGRISEAIIISAPRVFNRLVLCNRYFDCRPYVVGKPGCELPVAPVDPGTTTVPDRVLV 112
DQ 61 -----IISVVPPIMHAMETMCVRVFNIRPLVIG-PGIKTGLNKLKVDNPREIGSDRIV 112
QY 113 NTVAGYDRHGDLIVDFGTATTEDVVPADGAYIGGVIAAGVNLNLEALHMAAALPHVD 172
DQ 113 NAVAASEEYGPVIVDFGTATTTCYIDEGVYOGGAIAFGIMISTEALYNRAAKLPRVD 172
QY 173 VTKPGQVIGTNVACIQSGVYWGIVGIVRQIRMERDRPMKVIATGGLASFLDLGFD 232

Db 173 IAESSQIIIGKSTVSSMQAGIFGYFGQCEGIIAEMKKQSNTPSPVVATGGLARMITERKSS 232
QY 233 LFDKVEDDLTMHGLRLIFDYNK 254
DQ 233 AVDILDPLTLKGLLELYRRNK 254

RESULT 7
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C:Accession: S66100; E69740
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis ch
A:Reference number: S65967; MUID:96051385; PMID:7584024
A:Accession: S66100
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <OGA>
A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:dl005847; PID:g467

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69740
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <KUN>
A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:el1820
A:Experimental source: strain 168
C:Genetics:
A:Gene: yacB
A:Start codon: TTG
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 33.3%; Score 455; DB 2; Length 233;
Best Local Similarity 44.1%; Pred. No. 1.6e-32;
Matches 93; Conservative 38; Mismatches 78; Indels 2; Gaps 2;

QY 1 MLLCIDCGNTNTVFSVMDGTDFAATWRIATDHRRTADEYFWLNLMLQLKGLQ-GRISEA 59
DQ 1 MLLVIDGNTNTVLGVYHDKLEYHWRITSRKTEDEFGMLTSLRSLFHSGLMFEQIDGI 60
QY 60 IISSTAPRVFNRLVLCNRYFDCRPYVVGKPGCELPVAPVDPGTTTVPDRLVNTVAGYD 119
DQ 61 IISVVPPIMFALERMCTKYFHIEPIQVIG-PGKMTGLNLYDNPREVGADRVNVAATH 119
QY 120 RHGDLIVDFGTATTEDVVPADGAYIGGVIAAGVNLNLEALHMAAALPHVDVTKPGOV 179
DQ 120 LYGNPLIVDFGTATTTCYIDENKQYMGGAIPATGISTEALYSRAAKLPRIEIRPONI 179
QY 180 IGTNTVACIQSGVYWGIVGIVRQIRME 210
DQ 180 IGKNTVSAMQSGILFGYGVQVEGIVKRMKQ 210

RESULT 8

RESULT 9
H86937
Conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H86937
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho-
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MID:21128732; PMID:11234002
A;Accession: H86937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:gl3092576; PIDN:CAC29740.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0232
C;Superfamily: Streptomyces coelicolor hypothetical protein SCB94.31C

4:

RESULT 11

D/72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima

1

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: D72320
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: D72320
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <ARN>
A:CROSS-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g4981417
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0883
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 21.6%; Score 295.5; DB 2; Length 246;
Best Local Similarity 33.7%; Pred. No. 1.8e-18;
Matches 86; Conservative 48; Mismatches 96; Indels 25; Gaps 10;

QY 1 MLLCICGNTNTVFSVW-DGTDFAATWRIATDHRRTADEYFVWLTLM-----QLKGLQG 54
DB 1 MYLLVDGNTSHVSFSDTGKTF-RRWRLSTGVQTDELFSLHPLLGDMAMRKIGIG- 58
QY 55 RISEAIIISTAPRVVNLRLCNRYFDCRP-YVVGKPGCELPVAPRDPGTTVGPDRLVN 113
DB 59 -----VASVPTQNTVIERFSQKXPHISPIWVAKNGC---VKWVNKPSEVAGDVRAN 109
QY 114 TVAGYDRGGDLIVVDFGTATFDVAPDGAYIGVIAFGVNLSEALHMAAALPHVDV 173
DB 110 VVAFVKEYKNGIIMDMGTTATVDLVV-NGSEGGAILPGFFMVMVHSLFRGTAKPLVEV 168
QY 174 TKPGVIGTNTVACIOSGVYWGICLVGEGIVQIRMERDRPMKVATGGLASLFD- 231
DB 169 KPADFVGGKTEENIRLGVNGSVYALEGIIIGRIK-EVYGLDPLVLTGGQSKIYKDMIKH 227
QY 232 DLDFKVEDLTMHGL 246
DB 228 EIFD---EDLTIGV 239

RESULT 12
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C:Accession: D71326
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin- rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo- they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: D71326
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273 <COL>
A:CROSS-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0431
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 21.5%; Score 293.5; DB 2; Length 273;
Best Local Similarity 32.4%; Pred. No. 3e-18;
Matches 82; Conservative 42; Mismatches 118; Indels 11; Gaps 6;

QY 2 LLLCICGNTNTVFSVWDGTD-----FAATWRIATDHRRTADEYFVWLTLMQLKGLQRI 56
DB 1 MLLIDVGNHVVFGI-QGENGRCVRCVRELFRLAPDARKTQDEYSLIIHALCERAGV-GRA 58

QY 57 S--EAIISSTAPRVVNLRLCNRYFDCRPYVVGKGC- LPVAPRDPGTTVGPDRLVN 113
DB 59 SLRDAFISVVPVLTKTTADAVAQISGVQPVVFGWAYEHLVPRPEPRAEIGDVLVN 118
QY 114 TVAGYDRGGDLIVVDFGTATFDVAPDGAYIGVIAFGVNLSEALHMAAALPHVDV 173
DB 119 AVAAVHFHRSACVVVDCGTALTFTTAVDGTGLIQGVIAIPGLRTAVQSLHTGTQALPLVPL 178
QY 174 TKPGVIGTNTVACIOSGVYWGICLVGEGIVQIRMERDRPMKVATGGLASLFDLGFDP 233
DB 179 ALPDSVLGKDTTHAVQGVVGRGTLFVIRAMIAQCKELGCRCAAVITGSLSLFSEVD- 237
QY 234 FDKVEDLTMHGL 246
DB 238 FPPIDAQLTSL 250

RESULT 13
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: F70165
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh- son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; V- : Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70165
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <LSE>
A:CROSS-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g268- A:Experimental source: strain B31

Query Match 14.4%; Score 197.5; DB 2; Length 262;
Best Local Similarity 23.7%; Pred. No. 7.7e-10;
Matches 60; Conservative 55; Mismatches 123; Indels 15; Gaps 6;

QY 3 LCIDCGTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLTLMQLKGLQRISEALIS 62
DB 9 LIIDIGNTSIAPLKKQVNLFIKMKTNLMRLYDEVISFFE-----ENFDENVKVFIS 63
QY 63 STAPRVVNLRLCNRYFDCRPYVVG-KPGCELPVAPRDPGTTVGPDRLVNTVAGYDRH 121
DB 64 SVVPLNTEFKNVIFSFKKIKPLFTIGFDLNTDTFNPYKSKDFLLGSDVFAANLVAIENY 123
QY 122 G-GDLIVVDFGTATFDVAPDGAYIGVIAFGVNLSEALHMAAALPHVDVTKPGQVI 180
DB 124 SFENVLVVDLGTACTIFAVSRQDGLGGIINSGLINSLDNLAYLKKFPISIPNNLL 183
QY 181 GTNTVACIOSGVYWGICLVGEGIVQIRMERDRPMKVATG- LASFDFLGFDFDK 236
DB 184 ERTTSGSVNSGLFYQYKYLIEGVYRDIQKMYKKENLIITGNGADLLSLIEIEF-IFN- 241
QY 237 VEDDLTMHGLRLI 249
DB 242 --IHLTVGVRIL 252

RESULT 14
E70465
hypothetical protein aq_1924 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: E70465
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V. Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.48734 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-6

Perfect score: 1367

Sequence: 1 MLLCICGNTNTVFSVMDGT.....DLTGHGLRLIFDYNKGLGA 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB ID	Description
1	455	33.3	233	1 YACB_BACSU	P37564 bacillus su
2	134	9.8	267	1 BAF_BORPE	Q45338 bordetella
3	94.5	6.9	440	1 DCDA_STRCO	Q9zbb5 streptomyce
4	89.5	6.5	336	1 GPDA_STRCO	Q9zbs0 streptomyce
5	89.5	6.5	391	1 METK_CLOAB	Q97f85 clostridium
6	89	6.5	341	1 TDH_ECOLI	P07913 escherichia
7	88	6.4	788	1 PUR2_YARLI	Q99148 y bifunctio
8	86.5	6.3	225	1 CRP_MOUSE	P14847 mus musculus
9	86	6.3	411	1 CSD_AERPE	Q9yab6 aeropyrum p
10	85.5	6.3	398	1 METK_STRPY	Q99z77 streptococc
11	85	6.2	398	1 METK_STAAM	Q99t79 staphylococ
12	83	6.1	370	1 YDIK_ECOLI	P77175 escherichia
13	82.5	6.0	415	1 HUTI_RHIME	Q31196 rhizobium m
14	82.5	6.0	1888	1 CAIE_CHICK	P32018 gallus gall
15	82	6.0	265	1 Y309_HALN1	Q9hsb8 halobacteri
16	81	5.9	525	1 C824_SOYBN	Q49859 glycine max
17	80.5	5.9	395	1 METK_NEUCR	P48466 neurospora
18	80.5	5.9	396	1 METK_STRPN	Q97rn9 streptococc
19	80.5	5.9	439	1 SECY_SYN7	P31159 synecococc
20	80.5	5.9	588	1 C166_CHICK	P42292 gallus gall
21	80	5.9	284	1 LPXD_METCA	Q9aip8 methylococc
22	80	5.9	399	1 METK_BACHD	Q9k7g9 bacillus ha
23	80	5.9	407	1 HUTI_AGRRH	Q9kwe5 agrobacteri
24	79.5	5.8	470	1 LACG_STAUA	P11175 staphylococ
25	79	5.8	493	1 CSD_RAT	P64611 rattus norv
26	79	5.8	588	1 FTSI_ECOLI	Q04286 escherichia
27	78.5	5.7	400	1 METK_BACSU	P54419 bacillus su
28	78	5.7	320	1 K6P1_VIBCH	Q9knp2 vibrio chol
29	78	5.7	370	1 POTF_ECOLI	P31133 escherichia
30	78	5.7	473	1 YV33_MYCTU	Q06250 mycobacteri
31	78	5.7	1171	1 DPOL_HSVT1	Q9yus3 herpesvirus
32	78	5.7	1171	1 DPOL_HSVT2	Q9yus2 herpesvirus
33	77.5	5.7	468	1 LACG_STRMU	P50978 streptococc

ALIGNMENTS

RESULT 1

ID	YACB_BACSU	STANDARD;	PRT;	233 AA.
AC	P37564:			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical protein yacB.			
GN	YACB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=96051385; PubMed=7584024;			
RA	Ogasawara N., Nakai S., Yoshikawa H.;			
RT	"Systematic sequencing of the 180 kilobase region of the Bacillus			
RL	subtilis chromosome containing the replication origin.";			
RL	DNA Res. 1:1-14(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris H., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidos A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park A.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott S.M.,			
RA	Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivoita C., Roche E., Roche H., Rose M., Sadale Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takouchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,			
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus			
RT	subtilis.";			
RL	Nature 390:249-256(1997).			
CC	-1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.			

Q04733 nocardia la
Q9rzp0 deinococcus
Q26255 c.trifuncti
Q9ye36 brucella me
O66662 aquifex aeo
P76316 escherichia
Q9riy0 deinococcus
P13714 bacillus su
Q9cbr9 mycobacteri
Q9mup3 mesostigma
O83772 treponema p
P47770 staphylococ

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMBL; D26185; BAA05305.1; -;
 DR EMBL; Z99104; CAB11846.1; -;
 DR Subtilisin; BG10133; yacB.
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMS; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 26217 MW; AA956E732C15DF44 CRC64;

 Query Match 33.3%; Score 455; DB 1; Length 233;
 Best Local Similarity 44.1%; Pred. No. 1.8e-31;
 Matches 93; Conservative 38; Mismatches 78; Indels 2; Gaps 2;

 QY 1 MLLCIDGNTNTVFSWGDGTFAATWRIATDHRRTADEYFVWLTMLQKLGQ-GRISA 59
 DB 1 MLLVIDGNTNTVGLGYHDGKLEYHRIETSRKTEDEFGMLRSLFDHSGLMFEQIDGI 60
 QY 60 IISSTAPRVVNLRLVLCNRYFCRPVYVCKPGCELPVAPRVDPGTGTVGPDRLVNTVAGYD 119
 DB 61 IISVVPPIMFALERMCTKFIHIEFOIVG-PGKMGKLNKYDNPKVEGADRIVNAVAIH 119
 QY 120 RHGGDLIVDFGTATTFDVVAPDGAYIGGVVAPGNLSLEALHMAAALPHVDVTKPGCV 179
 DB 120 LYGNPLIVDFGTATTCYIDENKQVMGAIAPGITISTEALYSRAAKLPRIETRPDNI 179
 QY 180 IGTNTVACISQGVWGYGLVGEIVRQIRME 210
 DB 180 IGTNTVACISQGVWGYGLVGEIVRQIRME 210

 RESULT 2
 BAF_BORPE
 ID BAF_BORPE STANDARD; PRT; 267 AA.
 AC Q45338; Q45373;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Bvg accessory factor.
 GN BAF.
 Bortetella pertussis.
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 Bortetella.
 NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP504;
 RX MEDLINE=95325323; PubMed=7601846;
 RA Deshazer D., Wood G.E., Friedman R.L.;
 RT "Identification of a Bortetella pertussis regulatory factor required
 RT for transcription of the pertussis toxin operon in Escherichia
 RT coli.";
 RL J. Bacteriol. 177:3801-3807(1995).
 RN [2]
 RP SEQUENCE OF 1-38 FROM N.A.
 RC STRAIN=BP504;
 RA Wood G.E., Friedman R.L.;
 RT "Identification of a bifa homolog in Bortetella pertussis.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBS databases.
 RN [3]
 RP SEQUENCE OF 239-267 FROM N.A.
 RC STRAIN=BP536;
 RX MEDLINE=96419162; PubMed=8821935;
 RA Allen A.G., Maskell D.J.;
 RT "The identification, cloning and mutagenesis of a genetic locus

RT required for lipopolysaccharide biosynthesis in Bortetella
 RT pertussis.";
 RL Mol. Microbiol. 19:37-52(1996).
 CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
 CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
 CC RNA POLYMERASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12020; AAA75361.1; -;
 DR EMBL; AF016461; AAC68834.1; -;
 DR EMBL; X90711; CAA62242.1; -;
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc_factor; 1.
 KW Transcription regulation; Activator.
 SQ SEQUENCE 267 AA; 27845 MW; 338A615F67B57901 CRC64;

 Query Match 9.8%; Score 134; DB 1; Length 267;
 Best Local Similarity 28.6%; Pred. No. 0.00028;
 Matches 71; Conservative 27; Mismatches 96; Indels 54; Gaps 15;

 QY 1 MLLCIDGNTNTVFSWGDGTFAATWR-----TATDHRRTADEYFVWLTNL-----M 47
 DB 1 MLLIDSGNSRLKVG-WFDDPADQAAREPAPAFDN-LDLALGRWLATLPRRQALGV 58
 QY 48 QLKGLQGRISEAIIISSTAPRVVNLRLV-LCN-RYFDCRPVYV-KPGCELPVAPRVDPGT 104
 DB 59 NVAGLAR-GEAIAT-----LRAGGCDLRLRAQLANGLNGYRNP-----D 100
 QY 105 TVGPDR---LVNTVAGYDRHGGDLIVDPGTATTFDVVAPDGAYIGGVVAPGNLSLEAL 161
 DB 101 QLGADRWACWGVGLARQPSVHPPLLVASFGTATTLDTIGPDNVFPFGLILPGPAMRGAL 160
 QY 162 HMAAALPHVDVTKPGQVIG---TNTVACISQGVWGYGLVGEIVRQIRMERDR---P 214
 DB 161 AYTGAHLPLAD-----GLVADYPIDTHQIAAGIAAQAQNG---ALVRQWLACRQRYGQAP 212
 QY 215 MKVIATGG 222
 DB 213 EIVVAGG 220

 RESULT 3
 DCDA_STRCO
 ID DCDA_STRCO STANDARD; PRT; 440 AA.
 AC Q9ZBH5;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Diaminopimelate decarboxylase (EC 4.1.1.20) (DAP decarboxylase).
 GN LYSA OR SCO6438 OR SC9B5.05.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RL coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: Meso-2,6-diaminoheptanedioate - L-lysine +
 CO(2).
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- PATHWAY: Lysine biosynthesis; last step.
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF ORNITHINE, DAP, AND ARGININE
 DECARBOXYLASES.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL035206; CAA22747.1; -
 CC HSSP; P00860; 70DC.
 DR InterPro: IPR000183; Decarboxylase2.
 DR Pfam; PF0278; Orn_DAP_Arg_dec; 1.
 DR Pfam; PF02784; Orn_Arg_dec_N; 1.
 DR PRINTS; PR01179; OADACRXLASE.
 DR TIGRFAMs; TIGR01048; lysA; 1.
 DR PROSITE; PS00878; ODR_DC_2.1; 1.
 DR PROSITE; PS00879; ODR_DC_2.2; 1.
 DR Lysine biosynthesis; Lyase; decarboxylase; Pyridoxal phosphate;
 KW Complete proteome.
 FT BINDING 61
 FT SEQUENCE 440 AA; 47548 MW; 8DC03E8650DB3921 CRC64;
 SQ
 Query Match 6.9%; Score 94.5; DB 1; Length 440;
 Best Local Similarity 22.68; Pred. No. 1.1; Mismatches 69; Indels 89; Gaps 11;
 Matches 56; Conservative 34;
 QY 23 AATWRIATDHRRATDEYFVWMLTLMQLKGLQGRISEAI-ISSAPRVVFNRLVLCNRYFD 81
 DB 92 AAGRVGGD-----DEFIVTADLLNRSTLRRVELGIPVNAAGSPQLDQ----- 136
 QY 82 CRPVYVKGKCELPVAPRVDPGTVGPDRLVNTVAGYDRHG-----GDLIVVDFGTAT 134
 DB 137 -----VGRAAPGHPFVIRINPFGHGHRSKRTKTGGEHSKHGIWHELESIALVD----- 186
 QY 135 TEDVAPDGAIVGVIAPGVNLSLEALHMAAALPHVDVTKPGQVIGTNTVACIQSGVYV 194
 DB 187 -----RHGLDLVLGHM-----H-----IGSGVDY 205
 QY 195 GYI-GLVEGIVRQIRMERDRPMKVATGGSLASL-----FDLGFDFLFDKVEDDLTM 243
 DB 206 GHLESVCETWVKQVRM-AGRIDRAISAGGSLSVPTPGDPEIDPDYFELWDAARRELV 264
 QY 244 ---HGLRL 248
 DB 265 ELGHPVRL 272

RESULT 4
 GPDNA_STRCO STANDARD; PRT; 336 AA.
 ID GPDNA_STRCO
 AC Q92BS0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-
 GN dependent glycerol-3-phosphate dehydrogenase).
 GN GISA OR SC0559 OR SC7A1.03.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RX MEDLINE-21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P)(+) -
 CC glycerone phosphate + NAD(P)H.
 CC -1- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate
 CC formation.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
 CC DEHYDROGENASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL034447; CAA22402.1; -
 DR InterPro: IPR001652; NAD_Gly3P_dh.
 DR Pfam; PF01210; NAD_Gly3P_dh; 1.
 DR PRINTS; PR00077; GPDHNRGNASE.
 DR PRODOM; PD001649; NAD_Gly3P_dh; 1.
 DR PROSITE; PS00957; NAD_G3PDH; 1.
 DR Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.
 KW SEQUENCE 336 AA; 34688 MW; 4E7BF65F9C640544 CRC64;
 SQ
 Query Match 6.5%; Score 89.5; DB 1; Length 336;
 Best Local Similarity 27.9%; Pred. No. 2.1;
 Matches 50; Conservative 14; Mismatches 68; Indels 47; Gaps 9;
 QY 98 PRVDPGTVGPDRLVNTVAGYDRHGDDLIVVDFGTATTFVAPDPAVIGG---VIAPGV 154
 DB 96 PLIAPGV-----LVSLMKG-----VELSAMRMSEVIGDVAKVGAERTAVTGP 140
 QY 155 NLSLE-ALHMAAALP-----HVDVTKPGQVIGTNTVACIQSGVYVIG 198
 DB 141 NLAREIAARMPAAAVACPDVTAQRLQAACHTTPYFRP--YTNTDVVGCGLGAVKNVIG 198
 QY 199 LVEGIVRQIRMERDRPMKVATGGSLASLFDKVEDDLTMHGLRLFDNKGIG 257
 DB 199 LAVGIADGMGL-GDNARKGSLITRGLAETTRIGVAL---GADPLTFSGL-----AGIG 246

RESULT 5
 METK_CLOAB STANDARD; PRT; 391 AA.
 ID METK_CLOAB
 AC Q97F85;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
 GN adenosyltransferase) (AdoMet synthetase) (MAT).
 GN METK OR CAC2856.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.


```
Db 107 CRNTIGVGNRPGCAEYLYVPAFNAFKI--PDNISDDLAIAFPFGNAV-----HTALS 159
QY 136 FDVAPADGAYIGGYIAPCVNLSLEALHMAAALPHVDVTKPGVIGTNTVACIQSGVYWG 195
Db 160 FDLVGED-VLVSAGPGIGI-----MAAAVAKHV-----GARNVITDVNEVRL 201
QY 196 YIGLVEGIVRQIRMERDPMKVIATGGGLASLFDLGFDL-----FDKVEDDLTMHGLRL 248
Db 202 ELARKMGITRAVNVAKENLNDVMAELGTEGFDVGLMSGAPPARTMLDTMN-HGGRI 259

RESULT 7
PUR2_YARLI
ID PUR2_YARLI STANDARD; PRT; 788 AA.
AC Q99148;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bifunctional purine biosynthetic protein ADEL [Includes:
DE Phosphoribosylamine-glycine ligase (EC 6.3.4.13) (GARS) (Glycinamide
DE ribonucleotide synthetase) (Phosphoribosylglycinamide synthetase);
DE Phosphoribosylformylglycinamide phosphoribosyltransferase (EC 6.3.3.1) (AIRS)
DE (Phosphoribosyl-aminimidazole synthetase) (AIR synthase)].
GN ADEL.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA Strick C.A., James L.C., Cole K.E., Elsenboss L.A.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + 5-phospho-D-ribosylamine + glycine = ADP
CC + phosphate + N(1)-(5-phospho-D-ribosyl)glycinamide.
CC -!- CATALYTIC ACTIVITY: ATP + 2-(formamido)-N(1)-(5-phospho-D-
CC ribosyl)imidazole = ADP + phosphate + 5-amino-1-(5-phospho-D-
CC ribosyl)imidazole.
CC -!- PATHWAY: SECOND AND FIFTH STEPS IN DE NOVO PURINE BIOSYNTHESIS.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GARS FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE AIRS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; U40565; AAA85393.1; -
DR HSP; P15640; IGSO.
DR InterPro; IPR000728; AIRS_related.
DR InterPro; IPR000115; Gars.
DR InterPro; IPR004733; PurM_cligase.
DR Pfam; PF00586; AIRS; 1.
DR Pfam; PF01071; GARS; 1.
DR Pfam; PF02769; AIRS.C; 1.
DR Pfam; PF02842; GARS.B; 1.
DR Pfam; PF02843; GARS.C; 1.
DR Pfam; PF02844; GARS.N; 1.
DR TIGRFAMs; TIGR00877; purD; 1.
DR TIGRFAMs; TIGR00878; purM; 1.
DR PROSITE; PS00184; GARS; 1.
KW Multifunctional enzyme; Purine biosynthesis; Ligase.
FT DOMAIN 1 432
FT DOMAIN 442 752
FT SEQUENCE 788 AA; 83758 MW; 6EC170ECB27AFCF4 CRC64;

Query Match 6.4%; Score 88; DB 1; Length 788;
Best Local Similarity 26.5%; Pred. No. 7.2;
Matches 65; Conservative 27; Mismatches 81; Indels 72; Gaps 16;

QY 44 NTLMQLGLOGRISAEIISSTAPRV-----VFNLRLVLCNRYFDCRPVYWGKP---G 91
```

```
Db 452 NKLQV-----QIKEKVKSTARPGTSDSVIGGFGGLFDLKAAGFR-----DPLLVGATDGVG 501
QY 92 CELPVAPRVDPGTTVPDRLVNTVAGYDRHGGD-LIVVDFTGTTATFDVVPDCAVIGGYI 150
Db 502 TKLTKIAQSIDKHDVTGIDLVAMNVDLVVYVGAEPFLVDYATGKLDVNAA-AAFVGGV- 559
QY 151 APGVNLSLEALHMAAALPHVDVTKPGVIGTNTVACIQSGVYWG-----YIGLVEG 202
Db 560 -----ADGCIQACAL-----IGGETAE--MPGIYIYNDYDANGTSIGAVE- 598
QY 203 IVRQIRMERDPMKVIATG-----GLAS--LFDLGFDFLKVDEDDLTMHGLRLI----FDY 252
Db 599 --RDVAVLPR---MDEIAKGDAITGLASSGVHSGFSLVRKI---IEHAGLTVTDACPMQ 650
QY 253 NKG LG 257
Db 651 SKSLG 655

RESULT 8
CRP_MOUSE
ID CRP_MOUSE STANDARD; PRT; 225 AA.
AC P14847;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE C-reactive protein precursor.
GN PTX1 OR CRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLIN=89050112; PubMed=3190681;
RA Ohnishi S., Maeda S., Nishiguchi S., Arai T., Shimada K.;
RL "Structure of the mouse C-reactive protein gene.";
RL Biochem. Biophys. Res. Commun. 156:814-822(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBA/J; TISSUE=Liver;
RX MEDLIN=9019719; PubMed=2310378;
RA Whitehead A.S., Zahedi K., Rits M., Mortensen R.F., Lelias J.M.;
RL "Mouse C-reactive protein. Generation of cDNA clones, structural
RL analysis, and induction of mRNA during inflammation.";
RL Biochem. J. 266:283-290(1990).
CC -!- FUNCTION: CRP DISPLAYS SEVERAL FUNCTIONS ASSOCIATED WITH HOST
CC DEFENSE: IT PROMOTES AGGLUTINATION, BACTERIAL CAPSULAR SWELLING,
CC PHAGOCYTOSIS, & COMPLEMENT FIXATION THROUGH ITS CALCIUM-DEPENDENT
CC BINDING TO PHOSPHORYLCHOLINE.
CC -!- SUBUNIT: HOMOPENTAMER. PENTAXIN (OR PENTAXIN) HAVE A DISCOID
CC ARRANGEMENT OF 5 NONCOVALENTLY BOUND SUBUNITS.
CC -!- INDUCTION: THE CONCENTRATION OF CRP IN PLASMA INCREASES GREATLY
CC DURING ACUTE PHASE RESPONSE TO TISSUE INJURY OR INFLAMMATION.
CC -!- SIMILARITY: BELONGS TO THE PENTAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; X13588; CAA31928.1; -
DR EMBL; X17496; CAA35531.1; -
DR PIR; A31583; A31583.
DR PIR; S08286; S08286.
DR HSP; P02741; ICRV.
DR MGD; MGI:88512; Crp.
DR InterPro; IPR001759; Pentaxin.
DR Pfam; PF00354; pentaxin; 1.
```



```
KW Complete proteome.
FT NP_BIND 271 278 ATP (POTENTIAL).
FT METAL 18 18 MAGNESIUM (BY SIMILARITY).
FT METAL 283 283 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 398 AA; 43179 MW; E9CD98127BC959B CRC64;

Query Match 6.3%; Score 85.5; DB 1; Length 398;
Best Local Similarity 25.8%; Pred. No. 5.4;
Matches 42; Conservative 27; Mismatches 69; Indels 25; Gaps 7;

QY 85 YVVGKPGCELPVAPRDPGTVGPDRLVNTVAGYDRHGG-----DLIVDFGTATFDV 138
Db 242 FVIG-----GPGDSGLT-GRKIIIVDTYGGSRHGGAFSGKDATKVDRSAYARY 292
QY 139 VAPDAGYIGVIAAGVNLSEALHMAAALPHVDVTKPQGVIGNTV--ACIQSGVYWG 196
Db 293 IAKNLVAAGLVTKAEVQLAY-AIGVAPVSVRD-----TFGTSTPEAVLEAAROVF 345
QY 197 IGLVEGIVQRTMRDRPMKVIATGGGLASL-FDLGFDLFDKVE 238
Db 346 DLRPAGIIQMLDLKRPYKQTAAYGHMGRTDIDLPLWRLNKVD 388

RESULT 11
METK_STAAM STANDARD; PRT; 398 AA.
AC Q99T79;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
DE adenosyltransferase) (Adomet synthetase) (MAT).
GN METK OR SAV1790 OR SA1608.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashta A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RA "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
CC -|- FUNCTION: Catalyzes the formation of S-adenosylmethionine from
CC methionine and ATP. The overall synthetic reaction is composed of
CC two sequential steps, Adomet formation and the subsequent
CC triphosphate hydrolysis which occurs prior to release of
CC Adomet from the enzyme (By similarity).
CC -|- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC diphosphate + S-adenosyl-L-methionine.
CC -|- COFACTOR: 2 divalent ions, such as magnesium, and 1 potassium ion
CC bind per subunit (By similarity).
CC -|- PATHWAY: Activated methyl cycle.
CC -|- SUBUNIT: Homotetramer (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; AP003363; BAB57952.1; -.
DR EMBL; AP003135; BAB42876.1; -.
DR HSSP; P04384; LMXB.
DR InterPro; IPR002133; S-Adomet_synt.
DR Pfam; PF00438; S-Adomet_synt; 1.
DR Pfam; PF02772; S-Adomet_syntD2; 1.
DR Pfam; PF02773; S-Adomet_syntD3; 1.
DR TIGRFAMs; TIGR01034; metK; 1.
DR PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
DR PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
KW Transferase; One-carbon metabolism; ATP-binding; Magnesium;
KW Complete proteome.
FT NP_BIND 273 280 ATP (POTENTIAL).
FT METAL 19 19 MAGNESIUM (BY SIMILARITY).
FT METAL 285 285 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 398 AA; 43790 MW; 14E8C5C2C0E5BA07 CRC64;

Query Match 6.2%; Score 85; DB 1; Length 398;
Best Local Similarity 22.3%; Pred. No. 6;
Matches 50; Conservative 37; Mismatches 73; Indels 64; Gaps 13;

QY 28 IATDHRRTADEYFVWLNTLMQLKGLQGRISEALISSTAPRVVFNLRVLCNRYFDCRP--- 84
Db 199 VSTQH---ADD-----VTLEQIQEDIKAHVIYTPENLINEQT-----KEYINPTGR 243
QY 85 YVVGKPGCELPVAPRDPGTVGPDRLVNTVAGYDRHGG-----DLIVDFGTATFDV 138
Db 244 FVIG-----GPGDAGLT-GRKIIIVDTYGGYARHGGCGCFKDPKVDRSAAAYARY 294
QY 139 VAPD-----GAYIGGVIAAGVNLSEAL-----HMAAALPHVDVTKPQ 177
Db 295 VAKNIVAAGLADQCEQLAYAGVAEP-VSIAIDTFGTGKVGSEGLVAVRKHFDL-RPA 352
QY 178 GVIGTNTVACIQSGVY-----WGYIGLVEGIVRQIRMERDRPMK 216
Db 353 GII---KMLDLKQPIYKQTAAYGHFGRTDVLFPWEKLDKVEELK 393

RESULT 12
YDIK_ECOLI STANDARD; PRT; 370 AA.
AC P77175;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydiK.
GN YDIK OR B1688.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
```

DNA Res. 3:363-377(1996).
 RL -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0118 (PERM) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000264; AAC74758.1; -
 CC EMBL; D90811; BAA15450.1; -
 CC EcoGene; EGI3970; ydik.
 DR InterPro: IPR002549; UPF0118.
 DR Pfam: PF01594; UPF0118; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 18 38 POTENTIAL.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 249 269 POTENTIAL.
 FT TRANSMEM 275 295 POTENTIAL.
 FT TRANSMEM 309 329 POTENTIAL.
 SQ SEQUENCE 370 AA; 39841 MW; 14B7AD79D8C56248 CRC64;

 Query Match 6.1%; Score 83; DB 1; Length 370;
 Best Local Similarity 20.9%; Pred. No. 8.1;
 Matches 65; Conservative 37; Mismatches 85; Indels 124; Gaps 16;

 QY 17 WDCGTPAAWRIATDHRRTADEYFVWNLTMOLKGLQ-GRISEAIISSAPRV---VFNL 72
 DB 40 WAGTVVIAWTPV-----LLRLQKIFGRSLAVLWTLVWTFIIFI 82
 QY 73 RVLNRYFDCRPYVVGKPCGCELPVAPRVDPGTVGPD-RLVNTV-----AGYDR--- 120
 DB 83 ALLVNS-----IVDGS-PLIKAISSGDMTLPDLAWLNTIPVIGAKLYAGWNLDD 132
 QY 121 HGGDLIVD---FGTATFDV--VAPDCAIY-----CGVTAAGVN--- 155
 DB 133 MGTALMAKVPYIGTITTFWFGQAAHIGFRVHCALMLLFALLYWRGEVQAQGIHFPA 192
 QY 156 -----LSLEALHMAAALPHVDVTKPGVGIGTNTVA----- 186
 DB 193 TRLAGVRGDAVLLAAQAIRAVAGLVVVTALVQAVLGGIGLAVSGVPYATLITVLMILSC 252
 187 CIQSG-----VYW-----GYIGLVEGIVRQ--IRMERDRPMKVIA 219
 DB 253 LVQLGPLPLIPAIILWYGTGTWGTGTVLLVNSGVVGTLDNVRPMLIRMGADLPILIL 312
 QY 220 TGGLASLFDLG 230
 DB 313 SGVIGGLIAFG 323

 RESULT 13
 HUTI_RHIME STANDARD; PRT; 415 AA.
 AC O31196;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Imidazolonepropionase (EC 3.5.2.7) (Imidazolone-5-propionate
 DE hydrolase).
 DE HUTI OR RB0832 OR SBE21166.
 GN Rhizobium meliloti (Sinorhizobium meliloti).
 OS Plasmid pSymb (megaplasmid 2).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;

SEQUENCE FROM N.A.
 RP STRAIN=1021;
 RX MEDLINE=21396508; PubMed=11481431;
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
 RA Golding B., Puehler A.;
 RT "The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliloti".
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 RN [2]
 SEQUENCE OF 11-415 FROM N.A.
 RP STRAIN=RCR2011 / SU47;
 RC Uhde C., Schmidt R., Droege M., Jording D., Puehler A.,
 RA Selbitschka W.;
 RT "A Sinorhizobium meliloti hutH-mutant, defective in the histidine
 RT degrading enzyme histidase, is impaired in stationary phase
 RT survival".
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 4-Imidazolone-5-propanoate + H(2)O = N-
 CC formimino-L-glutamate.
 CC -1- PATHWAY: Histidine degradation; third step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE HUTI FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINUS
 CC DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL603645; CAC49232.1; ALT_INIT.
 DR EMBL; AF032903; AAB86962.1; ALT_FRAME.
 DR InterPro: IPR002604; ATZ_TRZ.
 DR Pfam: PF01685; ATZ_TRZ; 1.
 DR TIGRfam: TIGR01224; hutI; 1.
 KW Hydrolase; Histidine metabolism; Plasmid; Complete proteome.
 SQ SEQUENCE 415 AA; 44058 MW; 7B131BB00D501A13 CRC64;

 Query Match 6.0%; Score 82.5; DB 1; Length 415;
 Best Local Similarity 24.6%; Pred. No. 10;
 Matches 50; Conservative 32; Mismatches 62; Indels 59; Gaps 13;

 QY 41 VMLN---TLMQLKGLQGRISAIISSTAPRVF-----NLRVLCNRYFDCRPYVVG 88
 DB 14 LWRNARLATREELPLGIIEDGVIAVRGIRIVYAGPEAGLPSELARADQVDFCE----- 68
 QY 89 KPGCELPVAPRVDPGTVGP---DRLVNTVAGVDRHGGDLIVVDFG---TATFDVVAAPD 142
 DB 69 -----GRWTPALIDCHTHIV-----HGGNR-AREFOURLEGATYEIARA 108
 QY 143 GAYIGGVTAAGVYNLSLEALHMAAALPHVDVTKPGVIGTNTVACIQSGVYIGL-VE 201
 DB 109 GGGIASTVEATNALSVEAL--VEAALPLDITLLAEGV---STVE-VKSG-----YGLNVE 157
 QY 202 GIVRQI-----RMERDRPMKVIAV 220
 DB 158 AELKMLRAARLESRLRPVIVTS 180

 RESULT 14
 CALE_CHICK STANDARD; PRT; 1888 AA.
 ID CALE_CHICK
 AC P32018;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha I(XIV) chain precursor (Undulin).
 GN COL14A1.

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 21.0426 seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-6
Perfect score: 1367
Sequence: 1 MLLCIDCGNTNTVFSVWDGT.....DLTMMHGLRLIFDYNKGLGA 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_21.*
- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertibrate.*
- 14: sp.unclassified.*
- 15: sp.rvirus.*
- 16: sp.bacteriap.*
- 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	658	48.1	261	16 Q9A6Z1	Q9A6Z1 caulobacter
2	537	39.3	258	2 Q9F9B5	Q9F9B5 bacillus st
3	528	38.6	254	16 Q9RGH5	Q9RGH5 bacillus ha
4	527	38.6	265	16 Q9X8N6	Q9X8N6 streptomyce
5	519	38.0	255	16 Q8R7M2	Q8R7M2 thermoanaer
6	495	36.2	259	16 Q8XHL5	Q8XHL5 clostridium
7	485	35.5	273	16 Q97EB4	Q97EB4 clostridium
8	480	35.1	259	16 Q8YAC5	Q8YAC5 listeria in
9	479	35.0	259	16 Q92F54	Q92F54 listeria in
10	431.5	31.6	272	16 Q06282	Q06282 mycobacteri
11	431.5	31.6	274	16 Q9CD56	Q9CD56 mycobacteri
12	387	28.3	256	16 Q8RFE4	Q8RFE4 mycobacteri
13	347.5	25.4	262	16 Q9RX54	Q9RX54 deinococcus
14	295.5	21.6	246	16 Q9WZY5	Q9WZY5 thermotoga
15	293.5	21.5	273	16 Q83446	Q83446 treponema p
16	240.5	17.6	212	2 Q32514	Q32514 desulfovibr

17	197.5	14.4	262	16	O51477	borrelia bu
18	178	13.0	295	16	O8Y2M4	ralstonia s
19	156.5	11.6	229	16	O67753	aquifex ae
20	140	10.2	592	16	Q9JW17	neisseria m
21	136	9.9	248	16	Q9HWC1	pseudomonas
22	134	9.8	592	16	Q9JXF1	neisseria m
23	129	9.4	276	16	O8YGD7	anabaena sp
24	125.5	9.2	242	16	Q9PC14	xyella fas
25	117	8.6	257	16	P7A045	synecocyst
26	96	7.0	209	16	Q9PIA9	campylobact
27	95	6.9	448	16	Q9L2L6	streptomyce
28	93.5	6.8	225	11	Q91XB3	mus musculus
29	92	6.7	334	17	Q9UZJ4	pyrococcus
30	91.5	6.7	691	5	Q95R88	pyrococcus
31	91.5	6.7	704	5	Q9VVT8	drosophila
32	90	6.6	56	2	P94305	bacillus ps
33	89.5	6.5	391	16	Q97F85	clostridium
34	89	6.5	341	16	O8XEJ1	escherichia
35	87	6.4	425	16	Q98NK4	rhizobium l
36	86.5	6.3	3686	2	Q93H58	streptomyce
37	86	6.3	341	16	O8XIE4	clostridium
38	85.5	6.3	356	16	Q92W80	rhizobium m
39	85.5	6.3	398	16	Q99Z77	streptococc
40	85	6.2	344	16	Q9PDK8	xyella fas
41	85	6.2	383	2	Q9LCS7	streptomyce
42	85	6.2	398	16	Q99T79	staphylococ
43	85	6.2	6781	12	Q91AV2	porcine epi
44	84.5	6.2	299	16	Q9RW02	deinococcus
45	84.5	6.2	341	16	Q8ZL52	salmonella

ALIGNMENTS

RESULT 1

Q9A6Z1	PRELIMINARY;	PRT;	261 AA.
ID	Q9A6Z1		
AC	Q9A6Z1:		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DE	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Transcriptional activator, putative, Baf family.		
CC	CC1935.		
OS	Caulobacter crescentus.		
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;		
OC	Caulobacter.		
OX	NCBI_taxid=155892;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-ATCC 19089 / CB15;		
RX	MEDLINE-21173698; PubMed-11259647;		
RA	Nierman W.C., Feidblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,		
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,		
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,		
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,		
RA	Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,		
RA	Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,		
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;		
RT	"Complete genome sequence of Caulobacter crescentus.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).		
DR	EMBL; AE005867; AAK23910.1;		
DR	TIGR; CC1935;		
DR	InterPro; IPR004619; Baf.		
DR	Pfam; PF03309; Bvg_acc_factor; 1.		
DR	TIGRFAMS; TIGR00671; baf; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;		
Query Match	48.1%;	Score	658; DB 16; Length 261;
Best Local Similarity	50.8%;	Pred. No.	7.5e-48;
Matches 130;	Conservative	48;	Mismatches 74; Indels 3;
QY	1 MLLCIDCGNTNTVFSVWDGTDFATWRTATDHRRTADEYFWLNTLMQLKGLGR-ISEA 59		

RX	MEDLINE=21992816; PubMed=11997336;
RA	Bao Q., Tian Y., Li W., Xuan Z., Ru S., Dong W., Yang J.,
RA	Chen Y., Xue Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;
RT	"A complete sequence of T. tengcongensis genome.";
RL	Genome Res 12:689-700(2002).
DR	EMBL: AB013180; AAM25520.1; -.
KW	Complete proteome; Hypothetical protein.
SQ	SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;
 Query Match 38.0%; Score 519; DB 16; Length 255;	
Best Local Similarity 42.7%; Pred.No. 4.3e-36;	
Matches 109; Conservative 51; Mismatches 93; Indels 2; Gaps 2;	
QY	1 MLICIDCGNTNTVFSVMDGTFDAATWRIATDHRHTADEYFVWLNTLMQLKGLO-GRISEA 59 : : : : : : : : : : : : : : : : : : : :
Db	1 MLAFPDVGNITVMGVFKGKLLHSFRISTDKNKTVDGYGLMVLNQILGYNGISLITEIDVV 60 : : : : : : : : : : : : : : : : : : : :
QY	60 IISSTAPRVFNLRVLCNVFCRPPYVGKPGCELVPAPRPVDTGGTTVGPDLRVNTVAGDY 119 : : : : : : : : : : : : : : : : : : : : : : : : :
Db	61 IISSVYPPLMTLVQNLSKYFRFKPIWG-PGIKTGTINIKYDNPKPEYGADRIYNAAVE 119 : : : : : : : : : : : : : : : : : : : : : : : : :
QY	120 RHGGDLLIVDFGTFATTFDVVADPGAYIGGVIAPGVNLSEALHMAAALPHVDVTKPQG 179 : : : : : : : : : : : : : : : : : : : : : : : : :
Db	120 LYGGPVIVIDFGTATTCALSEKETGYGGIAPGLMISADALFORTAKLPKIDLTTPPTV 179 : : : : : : : : : : : : : : : : : : : : : : : : :
QY	180 IGNTVACIQSGVYGVYIGIVEGIVQRIMERDRPMKVATGGLASLFDLGFLDFOKVED 239 : : : : : : : : : : : : : : : : : : : : : : : : :
Db	180 INENTVASMQSIIIGHVGMVDIVTRMGEPAPSAYVATGTFANNIAEESKTIDTVNE 239 : : : : : : : : : : : : : : : : : : : : : : : : :
QY	240 DLTMHGLRLIFDYNK 254 : : : : :
Db	240 MLTLEGLRIYYERNK 254 : : : : :
 RESULT 6	
QBXHL5	PRELIMINARY; PRT; 259 AA.
ID	QBXHL5;
AC	08XHL5;
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-JUN-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Hypothetical protein CPE2468.
GN	CPE2468.
OS	Clostridium perfringens.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC	Clostridiales; Clostridiaceae; Clostridium.
ON	NCBL_TaxId=1502;
OX	[1]
RX	SEQUENCE FROM N.A.
RP	STRAIN=J3 / TYPE A;
RC	PUBMED=11792842;
RX	Shimizu T., Ohcawari K., Hirakawa H., Ohshima K., Yamashita A.,
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT	"Complete genome sequence of Clostridium perfringens, an anaerobic
RT	flesh-eater."
RT	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RL	ENBL: AP003194; BAB82174.1; -.
DR	InterPro: IPR004619; Baf.
DR	InterPro: IPR000515; BPD_transp.
DR	Pfam: PF03309; Bvg_acc_factor; 1.
DR	TIGREMS; TIGR00671; baf; 1.
DR	PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 259 AA; 28819 MW; ODSPFA3B7A145E10 CRC64;
 Query Match 36.2%; Score 495; DB 16; Length 259;	
Best Local Similarity 40.2%; Pred. No. 4.8e-34;	
Matches 104; Conservative 55; Mismatches 94; Indels 6; Gaps 3;	
QY	1 MLICIDCGNTNTVFSVMDGTFDAATWRIATDHRHTADEYFVWLNTLM-OLKGLQGRISEA 59 : : : : : : : : : : : : : : : : : : : :
Db	1 MILLIDVGNITVILGHNDKEKIASRWSDTSKTSDEYSIOYWOLFNOAKLPEDVEGI 60 : : : : : : : : : : : : : : : : : : : :

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein lin0253.
 GN LIN0253.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 OC NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / SEROVAR 6A;
 RX PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Gantier K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluster T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RT "Comparative genomics of Listeria species";
 RL Science 294:849-852(2001).
 DR EMBL; AL596164; CAC95486.1; -;
 DR ListiList; LIN00253; -;
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc.factor; 1.
 DR TIGRFAMS; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;
 Query Match 35.0%; Score 479; DB 16; Length 259;
 Best Local Similarity 39.3%; Pred. No. 1.le-32;
 Matches 103; Conservative 50; Mismatches 93; Indels 16; Gaps 4;
 QY 1 MLLCIDGNTNTVFSVMDGTDFAATWRIATDHRRTADE-----YFWLNL-TLMQLKGL 52
 DB 1 MILVIDGNTNCTGVYKEQKLLRHRWTTDRHTSDGLNLTNFSYANLTSDIQI 60
 QY 53 QGRISALISTAPRVFNVLNRYFDCRPYVVGKPCGELVAPRVDPGTTVGPDRLV 112
 DB 61 -----IISSVFPPIHMETMCRYFNIRPLVIG-PGIRTKLNKLVNPREIGSDRV 112
 QY 113 NTVAGYDRHGDDLIVDFGTATTFDVVAPDGYIGGVAPGVNLSLEALHMAAALPHVD 172
 DB 113 NAVAASEEIGTPVIVDFGTATTCYIDRAGVYGGGAIAPGIMISTEALYNRAAKLPVD 172
 QY 173 VTKPQGVIGTNTVACIQSVYWGVIIGLVEGIVRQIRMERDRPMKVIATGGGLASLFDL 232
 DB 173 IAESSQIIKSTVASMAGIFGYFGICEGIIAEMKKQSNTPPVVATGGGLARMTKSS 232
 QY 233 LFDKVEDDLTMHGLRLIFDYNK 254
 DB 233 AVDILDPFLTKLGLLELYRRNK 254
 RESULT 10
 ID O06282 PRELIMINARY; PRT; 272 AA.
 AC O06282;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 29.3 kDa protein (transcriptional activator, putative,
 DE Baf family).
 DE RV3600C OR MFCY07H7B.22 OR MT3706.
 GN Mycobacterium tuberculosis.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=1773;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.F., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Feltwell T., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / OSHKOSH;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z95557; CAB08944.1; -;
 DR EMBL; AE007170; AAK48063.1; -;
 DR TIGR; MT3706; -;
 DR TubercuList; RV3600c; -;
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc.factor; 1.
 DR TIGRFAMS; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 272 AA; 29304 MW; 5D70E6EF0F09AC8B CRC64;
 Query Match 31.6%; Score 431.5; DB 16; Length 272;
 Best Local Similarity 37.6%; Pred. No. 1.2e-28;
 Matches 100; Conservative 46; Mismatches 101; Indels 19; Gaps 4;
 QY 1 MLLCIDGNTNTVFSVMDG-----TDFATWRIATDHRRTADEYFWLNTLMQLKGLQRI 56
 DB 1 MLLAIDVNTNTHVGLLSGMKEHAKVVOQWRIRTESEVTADE-----LALTIDGLIGED 54
 QY 57 SEAIIS-----STAPRVFNVLNRYFDCRPYVVGKPCGELVAPRVDPGTTVGPDRLV 112
 DB 55 SERLTGTAALSTVSVLHEVRLMDQYWPSPVHLIEFGVGTGIPLLVDNKEVGAADRV 114
 QY 113 NTVAGYDRHGDDLIVDFGTATTFDVVAPDGYIGGVAPGVNLSLEALHMAAALPHVD 172
 DB 115 NCLAAAYDRFRKAAIIVDFGSSICVDSVSAKGEFLGGAIPGVQVSSAAAAARSALRRVE 174
 QY 173 VTKPQGVIGTNTVACIQSVYWGVIIGLVEGIVRQIR-----NERDRPMKVIATGGGLASL 227
 DB 175 LARPRSVVGNKNTVECMQAGVFGAGLVGLVRIREDVSGFSDVDHVAIVATGHTAPLL 234
 QY 238 DLGFDLFDKVEDDLTMHGLRLIFDYN 253
 DB 235 LPELHTVDHYDQHLTLQGLRLVFFERN 260
 RESULT 11
 ID O9CD56 PRELIMINARY; PRT; 274 AA.
 AC O9CD56;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein ML0232.
 GN ML0232.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=1769;

```

[1]
RN  COMPLETE PROTEOME.
RP  SEQUENCE FROM N.A.
RC  STRAIN=TN;
RX  MEDLINE=21128732; PubMed=11234002;
RA  Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA  Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA  Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA  Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA  Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA  Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA  Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA  Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA  Barrell B.G.;
RT  "Massive gene decay in the leprosy bacillus.";
RL  Nature 409:1007-1011(2001).
DR  EMBL: AL583917; CAC29740.1; -.
DR  Leproma; ML0232; -.
DR  InterPro: IPR004619; Baf.
DR  Pfam: PF03309; Bvg_acc_factor; 1.
DR  TIGRFAMS; TIGR00671; baf; 1.
DR  Hypothetical protein; Complete proteome.
SEQUENCE 274 AA; 29421 MW; 1C2E7358DEC78765 CRC64;

Query Match 31.6%; Score 431.5; DB 16; Length 274;
Best Local Similarity 37.2%; Pred. No. 1.2e-28;
Matches 100; Conservative 53; Mismatches 95; Indels 21; Gaps 4;

QY  1 MLCIDCGNTNTVFSVWDGTFDFAA---TWRTATDHRRTADEYFVWNTLMOLKGLQG-- 54
DB  1 MLLAIDVRNTHVTGVLSSKEHAKVVOQWRINTESEVTADE-----LALLIDGLIGDD 54

QY  55 --RISEAIISSTAPRVFNLRLCNRYFDCRPYVVGKPGCELPVAPRVPDGTGTVGDRIV 112
DB  55 SERLAGAALSTVPSVLHEVRIMLDQWPSVPHVLEPCVTRGIPLLVDNPREVGADRV 114

QY  113 NTVAGYDRUGGLIVVDFTATTDFVADPDGAYIGGVITAPGVNLSLEALHMAAALPHVD 172
DB  115 NCLAAFPKFGQAIVVDFGSSICDVVVSAGKEFLGGAIPAGVGVSSDAARSAALRRVE 174

QY  173 VTKPGQVIGTNTVACIQSGVYGYIGLVEGIVRQRMERDRPMKVIATGGLASLDGLDFD 225
DB  175 LARPSRVGKNTVECHQAGVVFAGLVGLVGRMQDVEEFGDGLGNRVAVVATGHTAP 234

QY  226 LFDLGFDFDKVEDDLTMHGLRLIFDYNK 254
DB  235 LLLPELHTVDHYDRHLTHLGLRLVFERNR 263

[1]
RN  COMPLETE PROTEOME.
RP  SEQUENCE FROM N.A.
RC  STRAIN=FE4
RX  MEDLINE=211889109; PubMed=11889109;
RA  Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA  Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA  Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA  Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA  Fongstein M., Kvirkides N., Overbeek R.;
RT  "Genome sequence and analysis of the oral bacterium Fusobacterium
RT  nucleatum strain ATCC 25586.";
RL  J. Bacteriol. 184:2005-2018(2002).
DR  EMBL: AE010586; AAL94957.1; -.

[1]
RN  COMPLETE PROTEOME.
RP  SEQUENCE FROM N.A.
RC  STRAIN=FE4
RX  MEDLINE=211889109; PubMed=10567266;
RA  White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA  Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA  Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA  Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA  Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA  Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA  Fraser C.M.;
RT  "Genome sequence of the radioresistant bacterium Deinococcus
RT  radiodurans R1.";
RL  Science 286:1571-1577(1999).
DR  EMBL: AE001905; AAF10040.1; -.
DR  TIGR: DR0461; -.
DR  InterPro: IPR004619; Baf.
DR  InterPro: IPR001220; Lectin_legB.
DR  Pfam: PF03309; Bvg_acc_factor; 1.
DR  TIGRFAMS; TIGR00671; baf; 1.
DR  PROSITE: PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
DR  Hypothetical protein; Complete proteome.
SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 25.4%; Score 347.5; DB 16; Length 262;
Best Local Similarity 36.0%; Pred. No. 1.5e-21;
Matches 90; Conservative 42; Mismatches 109; Indels 9; Gaps 6;

QY  2 LLCIDCGNTNTVFSVWDGTDFAATWRTATDHRRTADEYFVWNTLMOLKGLQGRISEAI 60
DB  6 LLAVDIGNTTIVLGLADASGALHTWTRINTREMLPDDLLAQLHGLTLAGAP-IPRAAV 64
```

```

KW  Complete proteome.
SQ  SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 28.3%; Score 387; DB 16; Length 256;
Best Local Similarity 32.8%; Pred. No. 6.6e-25;
Matches 85; Conservative 63; Mismatches 103; Indels 8; Gaps 7;

QY  1 MLCIDCGNTNTVFSVWDGTDFAATWRTATDHRRTADEYFVWNTLMOLKGLQ-GRISE 58
DB  1 MIIGIDIGNTHIVTGIYDNGELISFFRIATKDKTEDEYFYNNTKYNEISIKKYDA 60

QY  59 AIIISTAPRVFNLRLCNRYFDCRPYVVGKPGCELP--VAPRVPDGTGTVGDRIVNTVA 116
DB  61 ILISSVWPNIITITFOFFARKYKVEATIVDLEK-KLPFTFAKGIN-YTGFADRIIDITE 118

QY  117 GYDRH-GRGLIVVDFTATTDFVADPDGAYIGGVITAPGVNLSLEALHMAAALPHVDVTK 175
DB  119 AMQYPDKNLVIFDFGTATYDVL-KKGVIYGGGLPGIDMSINALYGTAKLPRVKFTT 177

QY  176 POGVIGTNTVACIQSGVYGYIGLVEGIVRQRMERDRPMKVIATGGLASLDGLDFD 235
DB  178 PSSVLGTDPMKIQIAIFPGYAGQIKHIKKINEELNEEIFVLATGGGLKILSAEIDEID 237

QY  236 KVEDDLTMHGLRLIFDYNK 254
DB  238 EYDANLSLKGTLTYKLNK 256
```

RESULT 13

```

Q9RX54
ID  Q9RX54      PRELIMINARY;      PRT;      262 AA.
AC  Q9RX54;
DT  01-WAY-2000 (TREMBLrel. 13, Created)
DT  01-WAY-2000 (TREMBLrel. 13, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Hypothetical protein DR0461.
GN  DR0461.
OS  Deinococcus radiodurans.
OC  Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC  Deinococcaceae; Deinococcus.
OX  NCBI_TaxID=1299;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=R1;
RX  MEDLINE=20036896; PubMed=10567266;
RA  White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA  Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA  Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA  Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA  Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA  Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA  Fraser C.M.;
RT  "Genome sequence of the radioresistant bacterium Deinococcus
RT  radiodurans R1.";
RL  Science 286:1571-1577(1999).
DR  EMBL: AE001905; AAF10040.1; -.
DR  TIGR: DR0461; -.
DR  InterPro: IPR004619; Baf.
DR  InterPro: IPR001220; Lectin_legB.
DR  Pfam: PF03309; Bvg_acc_factor; 1.
DR  TIGRFAMS; TIGR00671; baf; 1.
DR  PROSITE: PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
DR  Hypothetical protein; Complete proteome.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;
```

```

Query Match 25.4%; Score 347.5; DB 16; Length 262;
Best Local Similarity 36.0%; Pred. No. 1.5e-21;
Matches 90; Conservative 42; Mismatches 109; Indels 9; Gaps 6;

QY  2 LLCIDCGNTNTVFSVWDGTDFAATWRTATDHRRTADEYFVWNTLMOLKGLQGRISEAI 60
DB  6 LLAVDIGNTTIVLGLADASGALHTWTRINTREMLPDDLLAQLHGLTLAGAP-IPRAAV 64
```

```
QY 61 ISSTAPRVVFNLRVLCNRYFDCRPVYVKGPGCELP-VAPRVDPGTTPGDRLVNTVAGYD 119
Db 65 LSSVAPPVGENVALAKRHFMDAFVSAE--NLPDVTVELDTPGSVAGDRLCN-LFGAE 121
QY 120 RHGGDL---IYVDEGTATTFDVAPDGAIGVIGVAPGVNLSLEALHMAAALPHVDVTKP 176
Db 122 KYLGGLDYAVVVDGTSTNFDVGRGRLGILATGAQVSADALFAFAAKLPRTITLQAP 181
QY 177 QGVGTNTVACIQSGVYWGXYIGLVGIVQIRMERDRPMKVIATGGGLASLFDLGFDFLDFK 236
Db 182 ETAIGKNTVHALQSLVGFYAEWVDGLLRIRAEPLGAEVAVATGGFSRTVOGICQEIYD 241
QY 237 VEDDLTMHGL 246
Db 242 YDETTLRLGL 251

RESULT 14
Q9WZY5 PRELIMINARY: PRT: 246 AA.
01-NOV-1999 (TremBLrel. 12, Created)
01-NOV-1999 (TremBLrel. 12, Last sequence update)
01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 393:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -.
TIGR; TM0883; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc.factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 21.6%; Score 295.5; DB 16; Length 246;
Best Local Similarity 33.7%; Pred. No. 3.5e-17;
Matches 86; Conservative 48; Mismatches 96; Indels 25; Gaps 10;
QY 1 MLLCIDGNTNTVFSVM-DGTFDAATWRIATHRRTADEYFVWLNTLM-----QLKGLQG 54
Db 1 MYLLVDGNTHSVFESITDGKTF--RRWRLSTGVQTEDELFSHLPLLGDAAMREIKIG- 58
QY 55 RISEALISSTAPRVVFNLRVLCNRYFDCRP-VYVKGPGCELPVAPRVDPGTTPGDRLVN 113
Db 59 -----VASVYPTQNTVIERFSQKFIHISPIWAKNGC-----VKNNVKNPSEVGADRVAN 109
QY 114 TVAGYDRHGGDLIVVDFGTATTFDVAPDGAIGVIGVAPGVNLSLEALHMAAALPHVDV 173
Db 110 VVAFVKEGKNGIIDMGATTTVDLVV-NGSYEGAILPGFFMMVHSLFRGTAKLPLEV 168
QY 174 TKPOGVGTNTVACIQSGVYWGXYIGLVGIVQIRMERDRPMKVIATGGGLASLFD--LGF 231
Db 169 KPADFVVGKDFEENIRLGVNNGSVYALEGIIGRIK-EVYGDLPVVLITGGQSKIVKDMIKH 227
QY 232 DLFDKVEDDLTMHGL 246
Db 228 EIFD---EDLTIKGV 239
```

RESULT 15

```
083446 PRELIMINARY: PRT: 273 AA.
AC 083446;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein TP0431.
GN TP0431.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
Hatch E., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC85417.1; -.
TIGR; TP0431; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc.factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 21.5%; Score 293.5; DB 16; Length 273;
Best Local Similarity 32.4%; Pred. No. 6e-17;
Matches 82; Conservative 42; Mismatches 118; Indels 11; Gaps 6;
QY 2 LLCIDCGNTNTVFSVWDGTD-----FAATWRIATHRRTADEYFVWLNTLMQLKLGQRI 56
Db 1 MLLIDVGNHSHVYVFGI-QGNGRGVRCVRELFRLAPDARKTQDEYSLLIHALCERAGV-GRA 58
QY 57 S--EALISSTAPRVVFNLRVLCNRYFDCRPVYVKGPGCE-LPVAPRVDPGTTPGDRLVN 113
Db 59 SLRDAFISVVVPTTKTTADAVAQISGVQPVVFGPWAYEHLFPRPEPVRAEIGTDLVAN 118
QY 114 TVAGYDRHGGDLIVVDFGTATTFDVAPDGAIGVIGVAPGVNLSLEALHMAAALPHVDV 173
Db 119 AVAAVYHPSACVVVDCGTALTFTAVDGTGLIQGVIAIAPGLRTAVQSLHTGTALPLVPL 178
QY 174 TKPOGVGTNTVACIQSGVYWGXYIGLVGIVQIRMERDRPMKVIATGGGLASLFDLGF 233
Db 179 ALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQQCKELGCRCAAVITGGLSLFSEYD- 237
QY 234 FDKVEDDLTMHGL 246
Db 238 FPPIDAQLTSLGL 250

Search completed: June 24, 2003, 21:59:19
Job time : 22.0426 secs
```

THIS PAGE BLANK (USP10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.53698 seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-6
Perfect score: 1367
Sequence: 1 MLLCIDGNTNTVFSVMDGT.....DLTTHGLRLIFDYNKGLGA 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	6.1	1891	2	US-08-804-227C-12
2	84	6.1	1891	2	US-08-804-198-6
3	83.5	6.1	396	4	US-09-273-686-2
4	83.5	6.1	514	4	US-09-134-001C-4347
5	80	5.9	659	4	US-08-894-818B-1
6	80	5.9	659	4	US-09-445-472-12
7	79.5	5.8	464	4	US-08-914-375C-58
8	79.5	5.8	731	2	US-08-911-364-1
9	79.5	5.8	733	4	US-08-464-700-2
10	79	5.8	506	4	US-09-117-250-1
11	77.5	5.7	820	1	US-08-291-896-2
12	77.5	5.7	820	2	US-08-485-278-2
13	76.5	5.6	532	3	US-08-948-564-10
14	75.5	5.5	467	4	US-08-914-375C-57
15	75.5	5.5	1215	4	US-09-134-001C-5319
16	75	5.5	249	4	US-09-372-422A-30
17	75	5.5	415	4	US-09-134-001C-5077
18	75	5.5	495	4	US-09-217-490-2
19	75	5.5	614	4	US-09-134-001C-3061
20	74.5	5.4	387	3	US-08-689-421-23
21	74.5	5.4	387	4	US-09-389-528-23
22	74.5	5.4	387	4	US-09-181-827A-23
23	74.5	5.4	516	3	US-08-689-421-29
24	74.5	5.4	516	4	US-09-389-528-29
25	74.5	5.4	516	4	US-09-181-827A-29
26	74	5.4	601	4	US-09-173-300-4
27	73	5.3	979	4	US-09-323-872A-55

28	72.5	5.3	3170	2	US-07-642-734C-5	Sequence 5, Appli
29	72.5	5.3	3170	3	US-08-439-009A-5	Sequence 5, Appli
30	72.5	5.3	6095	4	US-09-144-085-2	Sequence 2, Appli
31	72	5.3	529	4	US-09-615-192A-405	Sequence 405, App
32	71.5	5.2	333	2	US-08-896-083-2	Sequence 2, Appli
33	71.5	5.2	333	2	US-08-896-083-4	Sequence 4, Appli
34	71.5	5.2	333	3	US-09-110-910A-2	Sequence 2, Appli
35	71.5	5.2	333	3	US-09-110-910A-4	Sequence 4, Appli
36	71	5.2	537	1	US-08-472-028A-2	Sequence 2, Appli
37	71	5.2	537	2	US-08-808-931-2	Sequence 2, Appli
38	71	5.2	537	3	US-08-808-333-2	Sequence 2, Appli
39	71	5.2	537	3	US-09-050-603A-2	Sequence 2, Appli
40	71	5.2	537	3	US-09-102-420B-2	Sequence 2, Appli
41	71	5.2	537	4	US-09-071-296-2	Sequence 2, Appli
42	71	5.2	537	4	US-09-196-268-2	Sequence 2, Appli
43	71	5.2	537	4	US-09-015-683-2	Sequence 2, Appli
44	71	5.2	537	4	US-09-191-998-2	Sequence 2, Appli
45	71	5.2	537	4	US-09-497-698-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-12
; Sequence 12, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1891 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-227C-12

Query Match 6.1%; Score 84; DB 2; Length 1891;
Best Local Similarity 24.9%; Pred. No. 5;
Matches 44; Conservative 18; Mismatches 61; Indels 54; Gaps 7;
QY 88 GKPGCELPVPRVDPGTTVGPDRLVNTVAGYDRHGGLIWDVDFGTATTFDVFVAPDGAYIG 147
DB 1070 GSPRTETP-----DRGTLLAA--RLAELARSPEGLAGVLLLPDPSG-----GAAVA 1111
QY 148 GVAPGVN-----LSLEALHMAAALPHVDVTKPQGVIGTNTVACIQSGVYWG----- 195

Db 1112 G--HPGLDQGTAAVLLTIGALTDAAVRAPLVWVTRGAVAGSGEVCPCAVGARVWGLGRVA 1169
QY 196 -----YTLGEGIVRQIRMERDRPMKVIAATGGGLASLFDLGLDFDKVEDDLTHWG 245
Db 1170 ALEVPVQWGLVDVAVAGVREWRVGVVAGGG-----EDQVAVRG 1211

RESULT 2

US-08-804-198-6
; Sequence 6, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostek, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1891 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-198-6

Query Match 6.1%; Score 84; DB 2; Length 1891;
Best Local Similarity 24.9%; Pred. No. 5;
Matches 44; Conservative 18; Mismatches 61; Indels 54; Gaps 7;
QY 88 GKPGCCLPVPVRDCTTVPDLVNTVAGYDRHGDLIVDFGTATTFDVAPDCAIG 147
Db 1070 GSPRTETP-----DRGTAA--RLAELARSPGLAGVLLLPDSG-----GAAVA 1111
QY 148 GVIAQGVN-----LSLEALHMAAALPHVDVTKPGQVIGTNTVACIOGVTWG----- 195
Db 1112 G--HPGLDQGTAAVLLTIGALTDAAVRAPLVWVTRGAVAGSGEVCPCAVGARVWGLGRVA 1169
QY 196 -----YTLGEGIVRQIRMERDRPMKVIAATGGGLASLFDLGLDFDKVEDDLTHWG 245
Db 1170 ALEVPVQWGLVDVAVAGVREWRVGVVAGGG-----EDQVAVRG 1211

RESULT 3

US-09-273-686-2
; Sequence 2, Application US/09273686
; Patent No. 6228625
; GENERAL INFORMATION:
; APPLICANT: Zalacain, Magdalena

; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Brown, James
; APPLICANT: Ingraham, Karen, A.
; APPLICANT: Chalker, Allison F.
; APPLICANT: SO, Chi Y.
; APPLICANT: Holmes, David J.
; APPLICANT: Van Horn, Stephanie
; APPLICANT: Warren, Richard L.
; TITLE OF INVENTION: metK
; FILE REFERENCE: GM10176
; CURRENT APPLICATION NUMBER: US/09/273,686
; CURRENT FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 60/106,767
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-273-686-2

Query Match 6.1%; Score 83.5; DB 4; Length 396;
Best Local Similarity 24.9%; Pred. No. 0.52;
Matches 44; Conservative 26; Mismatches 54; Indels 53; Gaps 10;
QY 85 YVVGKPGCCLPVPVRDCTTVPDLVNTVAGYDRHGG-----DLIVVDFGTATTFDV 138
Db 243 FVIG-----GPGDSGLT-GRKIIIVDTYGYSHRGGGAFSGDKATKVDASAYARY 293
QY 139 VAPDGAYIGGVIAPOGVNLSLEALHMAAALPHVDVTKPGV-----IGTNTVACIOGVTW 194
Db 294 IAKN-----IVAAGLAKKAE-VQLAYA----IGVAQPVSVRIDTGTCTVAESQ----- 337
QY 195 GYIGLVE-----GIVRQIRMERDRPMKVIAATGGGLASLFDLGLDFDKVE 238
Db 338 -----LEKAAQIFDLRPAIIQMLDLKRPYRQTSAYGHMGRTDIDLPERLDKVD 389

RESULT 4

US-09-134-001C-4347
; Sequence 4347, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4347
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4347

Query Match 6.1%; Score 83.5; DB 4; Length 514;
Best Local Similarity 21.1%; Pred. No. 0.78;
Matches 53; Conservative 32; Mismatches 73; Indels 93; Gaps 11;
QY 53 QGRISEALISSTAPRVFNL-----RVLCNRYFDCRPYVVKP----- 90
Db 254 KGEAVTIGTSGAIRTVINQPKTDEKGRIFC-YILDKQYVIGGPNNGGVVLRWLRLDEI 312
QY 91 -GCELPVAPRVDPGTTVGPDRLVNVVAGYDRHGDLIVVDFGTATTFDVAPDCAIGV 149
Db 313 LASEVETAKRIG-----VDPYDVLTOIASRVKPGAGSLIFH-----PYLAGE 354

Qy	102	PGRTV	---G	PDRLVNTVAGVDRHGGD	LLIVDFGRATTFD	VVAPD	-----	142
Db	138	DATSV	QIGADT	VWNSL	-GYDGS	VVAIVDTG	---IDANHPDLKGVIGWYDAVNGRS	192
Qy	143	-----	GAYIG	VIAPGVNLSLEALHMAAALP	PHVDVTPQGVIGNTVACIOGS	VYVW	194	
Db	193	TPYD	DG	HGTHVAGIVAGTGS	VNSOYIGVAPGA	--KLGVKVLGADGSGSVSTIIAGVDW	250	

RESULT 6

US 03 443 412 12
; Sequence 12, Application US/09445472
; Patent No. 6358726

Query Match 5.9%; Score 80; DB 4; Length 659;

Best local similarity 22.00, PDC: NO: 2.0;
Matches 41; Conservative 32; Mismatches 69; Indels 38; Gaps 6;

QV 46 LMOLKGLGRISEAIISSSTAPRVFNLRLVLCNRYFDCRPY----VVGKPGCELPAVRVD 101

Db 78 VLRLMGAQVKYSYKIIIPAVAVKIKARDLLLIAGMIDTGYFGNTRVSGIKFIQEDYKQVOD 137

Qy 102 PGTV---GPDRLVNTVAGYDRHGGDLIVVDEFGTATTFDVVAPD----- 142

138 D⁸T⁸S⁸V⁸O⁸I⁸C⁸A⁸T⁸V⁸N⁸S⁸I⁸-C⁸Y⁸D⁸C⁸S⁸C⁸W⁸V⁸A⁸I⁸V⁸T⁸C⁸-----I⁸D⁸A⁸N⁸H⁸R⁸O⁸I⁸K⁸C⁸K⁸V⁸I⁸C⁸W⁸D⁸A⁸V⁸N⁸C⁸S⁸ 192

[illegible]

Qy 143 -----GAYIGGVIAPGVNLSLEALHMAAALPHVDVTKPQGVIGTNTVACIQSGVYW 194

103 mpvddocwctwiazctvazctcsvnsqv tcvazpca - wfvvuvvzicazdcscsvstt r acvaw 350

DU 193 IFIDQGRGITHVAGTAVAGIGSVNSQITGVAFGA - KLVGVKVLGGADGGSVSITLGGVDW 230

RESULT 7
US-00-014-37EC-50

US-08-914-375C-58
: Sequence 58, Application US/08914375C

; Patent No. 6377893 .

GENERAL INFORMATION:

APPLICANT: STEVEN A. BENNER

Applications of Protein Structure Predictions

NUMBER OF SEQUENCES: 74

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Steven A. Benner
STREET: 1501 NW 68th Terrace

STREET: 1501 NW 60TH TERRACE
CITY: Gainesville

STATE: FL

PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 93/10050
FILING DATE: 17-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/291,896
FILING DATE: 17-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,409
REFERENCE/DOCKET NUMBER: P58114NA
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 820 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-278-2

Query Match 5.7%; Score 77.5; DB 2; Length 820;
Best Local Similarity 20.1%; Pred. No. 7.6; Indels 79; Gaps 14;
Matches 57; Conservative 47; Mismatches 100; Indels 79; Gaps 14;
QY 3 LCIDGNTVFSVMDG-----TQFAATWRIA-----TDHRTA-----DEY 39
DB 47 LVLDAGYVTGAGTSGGAVLPSPLAAGRRITIERVLVDVQETDLRNGKFFPEVHEDA 106
QY 40 FVWLTLMQLKGLGRISAAIISSTAPRVVFNRLVLCNRYFDCRPYVVGKPGCELPAVR 99
DB 107 FQYLTMLIQ--RCFGFRRLMKPS-----LLAKYYDAK-----QNRISNL 145
QY 100 VPGTTVGPDRLVN-----TVAGYDRHGDLIVDFGTAT---TFDVAPGAYI- 146
DB 145 ADP--SLEDQVNNRSMRYVDAAAGV--IGFGWFIOYSGGAVYRTFDKMRDGVSIK 201
QY 147 -----GGVIAPGVNLSLEALHMAAALPHVDYTKPGQVIGTNTVACIQSGVYWGVLVE 201
DB 202 DFGAQQGILNDKNDAPTKSLHSFS-----VFVPEGVENTSLVLSRCGLY----GTGG 251
QY 202 GIVRQIRMERDPMKVIATGGASLFDGLDFDKVEDDLTMH 244
DB 252 GTIKQYDRGNHLVFNMPDGMMLSTLTI---MGNKSDDSVQGH 291

RESULT 13
US-08-948-564-10
Sequence 10, Application US/08948564
Patent No. 6121512
GENERAL INFORMATION:
APPLICANT: Siminszky, Balazs
APPLICANT: Dewey, Ralph E.
APPLICANT: Corbin, Frederick T.
TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
METHOD OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: PO Box 37428
CITY: Raleigh
STATE: No. 6121512th Carolina
COUNTRY: USA
ZIP: 27627
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,564
FILING DATE:

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-409
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-854-1400
TELEFAX: 919-854-1401
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-564-10
Query Match 5.6%; Score 76.5; DB 3; Length 532;
Best Local Similarity 20.8%; Pred. No. 5.1;
Matches 47; Conservative 26; Mismatches 68; Indels 85; Gaps 10;
QY 11 NTVFSVMDGTDEAATWRIATDHRTADEYFVWLTLMQLKGLGRIS-----57
DB 324 NLILAGSDTMTLSLTWLS-----LLLNHQMELKVKQDELDYIGDKRKYVESD 372
QY 58 -----EAIISST-----APRVVFNRLVLCNRYFDCRPYVVGKPGCELPAVR-----100
DB 373 ITKLVTQIAIVKTRMRLYPPSPPLITLRAAME---DC-----TFSGGYHIPAGTRLMVNAWK 425
QY 101 -----DPGTTVGPDLVNTVAGYDRHGDLIVDFGTATTFDVAPDAGYIGVIA 151
DB 426 IHRDGRVSDP--HDFKPGRFLTSKDVVDYKQNYELVPEFS-----GRRAC 470
QY 152 PCVNLSLEALHMAAALPH-----VDVTKPGQVIGTNTVA 186
DB 471 PCASLALRVVHLTMARLLHSFNVASPSNQVVDMTESIGL--TNLKA 514
RESULT 14
US-08-914-375C-57
Sequence 57, Application US/08914375C
Patent No. 6377893
GENERAL INFORMATION:
APPLICANT: Steven A. Benner
Applications of Protein Structure Predictions
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: 1501 NW 68th Terrace
CITY: Gainesville
STATE: FL
COUNTRY: United States
ZIP: 32605-4147
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/914,375C
FILING DATE: 19-Aug-1997
CLASSIFICATION: 702/20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352 392 7773
TELEFAX: 352 331 0462
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 467
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: amino acid
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
FEATURE:

Search completed: June 24, 2003, 21:49:18
Job time : 8.53698 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.7629 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-6

Perfect score: 1367

Sequence: 1 MLLCIDCGNTNTVFVWDGT.....DDLTMHGLRLIFDYNKGLGA 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1367	100.0	258	9	US-09-813-453A-6
2	658	48.1	260	9	US-09-813-453A-51
3	554	40.5	255	9	US-09-813-453A-7
4	542	39.6	258	9	US-09-813-453A-49
5	537	39.3	258	9	US-09-813-453A-2
6	536	39.2	262	9	US-09-813-453A-45
7	528	38.6	254	9	US-09-813-453A-47
8	527	38.6	265	9	US-09-813-453A-4
9	496.5	36.3	256	9	US-09-813-453A-55
10	455	33.3	233	9	US-09-813-453A-17
11	431.5	31.6	272	9	US-09-813-453A-5
12	428.5	31.3	250	9	US-09-813-453A-3
13	427.5	31.3	272	9	US-09-712-363-276
14	394	28.8	219	9	US-09-813-453A-57
15	347.5	25.4	262	9	US-09-813-453A-8
16	295.5	21.6	246	9	US-09-813-453A-9
17	293.5	21.5	273	9	US-09-813-453A-10
18	265	19.4	257	9	US-09-813-453A-53
19	240.5	17.6	212	9	US-09-813-453A-59

20	224	16.4	244	9	US-09-813-453A-41	Sequence 41, Appl
21	197.5	14.4	262	9	US-09-813-453A-11	Sequence 11, Appl
22	168.5	12.3	241	9	US-09-813-453A-63	Sequence 63, Appl
23	158.5	11.6	229	9	US-09-813-453A-12	Sequence 12, Appl
24	140	10.2	592	9	US-09-813-453A-22	Sequence 22, Appl
25	139	10.2	460	9	US-09-813-453A-39	Sequence 39, Appl
26	136	9.9	248	9	US-09-813-453A-20	Sequence 20, Appl
27	134	9.8	267	9	US-09-813-453A-15	Sequence 15, Appl
28	134	9.8	592	9	US-09-813-453A-43	Sequence 43, Appl
29	132	9.7	249	9	US-09-813-453A-70	Sequence 70, Appl
30	125.5	9.2	242	9	US-09-813-453A-65	Sequence 65, Appl
31	124	9.1	249	9	US-09-813-453A-61	Sequence 61, Appl
32	117	8.6	257	9	US-09-813-453A-13	Sequence 13, Appl
33	96	7.0	209	9	US-09-813-453A-21	Sequence 21, Appl
34	93	6.8	925	9	US-09-924-097-14	Sequence 14, Appl
35	84.5	6.2	255	10	US-09-847-637B-9	Sequence 9, Appl
36	83.5	6.1	386	10	US-09-815-242-4878	Sequence 4878, Ap
37	83.5	6.1	394	10	US-09-815-242-10766	Sequence 10766, A
38	83.5	6.1	396	10	US-09-815-242-13413	Sequence 13413, A
39	83.5	6.1	396	10	US-09-805-847-2	Sequence 2, Appl
40	83.5	6.1	397	10	US-09-815-242-13591	Sequence 13591, A
41	81.5	6.0	397	10	US-09-815-242-5571	Sequence 5571, Ap
42	81.5	6.0	431	10	US-09-815-242-12281	Sequence 12281, A
43	81.5	6.0	431	10	US-09-815-242-12856	Sequence 12856, A
44	81.5	6.0	1265	10	US-09-862-027-80	Sequence 80, Appl
45	81	5.9	396	9	US-10-285-074-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-813-453A-6
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OCGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 258
; TYPE: P3T
; ORGANISM: Rhodobacter capsulatus
; US-09-813-453A-6

Query Match	100.0%	Score 1367;	DB 9;	Length 258;
Best Local Similarity	100.0%	Pred. No. 9.4e-131;		
Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLLCIDCGNTNTVFVWDGTDFAAFWRIATDHRRTADEYFVWLNTLMQLKGLQGRISEAI	60	
DB	1	MLLCIDCGNTNTVFVWDGTDFAAFWRIATDHRRTADEYFVWLNTLMQLKGLQGRISEAI	60	
QY	51	ISSATPRVFNLRVLCNRYFDCRPVYVKGPCGCELPVAPRVDPTGTTVPDRLVNTVAGYDR	120	
DB	51	ISSATPRVFNLRVLCNRYFDCRPVYVKGPCGCELPVAPRVDPTGTTVPDRLVNTVAGYDR	120	
QY	121	HGGDLIVYDFGTATTFDVAADGAYIGGVAPGNLSLEALHMAAALPHVDVTPQGVY	180	
DB	121	HGGDLIVYDFGTATTFDVAADGAYIGGVAPGNLSLEALHMAAALPHVDVTPQGVY	180	
QY	131	GTNTVACIQSGVYWGYYIGLVREGIVRQIRMERDRPMKVITATGGLASFLDGLDFDLKVEDD	240	

Db 180 ICKNTVSAMQSGILYGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDP 239
QY 240 DLTMHGLRLIFDYN 253
Db 240 FLTLGGLLYEKN 253

RESULT 5

US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRF
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 39.3%; Score 537; DB 9; Length 258;
Best Local Similarity 43.9%; Pred. No. 1.6e-46;
Matches 112; Conservative 46; Mismatches 95; Indels 2; Gaps 2;
QY 1 MLLCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLNTLMQLKGQGR-ISEA 59
Db 1 LLLVIDGNTNVLGVHDGKLEWHRIETSRHKTDEFEGLTSLRSLFDHSGLMFQIDGI 60
QY 60 IISSTAPRVNLRVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYD 119
Db 61 IISVVPPIFMALERMCKYFHIEPIQVIG-PCMKTGLNPKYDNPKEVGADRVNVAIAH 119
QY 120 RHGGDLIVVDGCTATTEDVAPDGAIGGVIAAGVNLSEALHMAAALPHVDVTKPGOV 179
Db 120 LYGNPLIVVDGCTATTTCYINEEKHYMGVITPGIMISAEALYSRAKLPRIETKPSV 179
QY 180 IGTNTVACIQSGVYWGIGLVGIVQIRMRDRPMKVIATGGLASLFDLFDKVED 239
Db 180 ICKNTVSAMQSGILYGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDCIDVDP 239
QY 240 DLTMHGLRLIFDYN 254
Db 240 FLTLGGLLYERNR 254

RESULT 6

US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRF
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match 39.2%; Score 536; DB 9; Length 262;
Best Local Similarity 41.6%; Pred. No. 2.1e-46;
Matches 107; Conservative 55; Mismatches 93; Indels 2; Gaps 2;
QY 1 MLLCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLNTLMQLKGQGR-ISEA 59
Db 1 MIFVLDVGNNTNVLGVFEGLRQHWRMETDRHKTDEYGLVQLLEHGLSFEDEVKGI 60
QY 60 IISSTAPRVNLRVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYD 119
Db 61 IVSSVPPPIFMALERMCKYFKIKPLVVG-POIKTGLNPKYDNPKEVGADRVNVAAGIH 119
QY 120 RHGGDLIVVDGCTATTEDVAPDGAIGGVIAAGVNLSEALHMAAALPHVDVTKPGOV 179
Db 120 LYGSPLIVVDGCTATTTCYINEEKHYMGVITPGIMISAEALYSRAKLPRIETKPSV 179
QY 180 IGTNTVACIQSGVYWGIGLVGIVQIRMRDRPMKVIATGGLASLFDLFDKVED 239
Db 180 VGNKTVSAMQSGILYGVQVEGIVSRMKKEAKQBPVKVIATGGLAKLISESNVIDVDP 239
QY 240 DLTMHGLRLIFDYNKGL 256
Db 240 FLTLGGLYLYERNANL 256

RESULT 7

US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRF
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 38.6%; Score 528; DB 9; Length 254;
Best Local Similarity 41.3%; Pred. No. 1.3e-45;
Matches 105; Conservative 55; Mismatches 92; Indels 2; Gaps 2;
QY 1 MLLCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLNTLMQLKGQGR-ISEA 59
Db 1 MLVIDVGNNTNVLGVQDETILVHWRLATSKQKTEDEYAMTVRSFLFDHAGLQFQIDGI 60
QY 60 IISSTAPRVNLRVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYD 119
Db 61 VISSVPPPMFSLQMKCKYFHVTPMIIG-POIKTGLNPKYDNPKEVGADRVNVAIAE 119
QY 120 RHGGDLIVVDGCTATTEDVAPDGAIGGVIAAGVNLSEALHMAAALPHVDVTKPGOV 179
Db 120 LYGPALIVVDGCTATTTCYINEEKQYAGGVIAAGVIAAGVIAAGVIAAGVIAAGVIAAGV 179

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 28.1855 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453A-61
Perfect score: 1269
Sequence: 1 MILEDCGNSFKWRVHVA.....QARVVPDLVFGVGLAMACPLD 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1269	100.0	249	AAU91178	Pantothenate kinas
2	849	66.9	249	AAU91182	Pantothenate kinas
3	794.5	62.6	248	AAU91164	Pantothenate kinas
4	245	19.3	592	20 AAY38615	Neisseria meningit
5	245	19.3	592	21 AAY74912	Neisseria meningit
6	245	19.3	592	23 AAU91169	Pantothenate kinas
7	244.5	19.3	389	21 AAU74909	Neisseria meningit
8	243	19.1	455	21 AAY74910	Neisseria meningit
9	243	19.1	592	20 AAY38616	Neisseria meningit
10	243	19.1	592	21 AAY74913	Neisseria meningit

11	243	19.1	592	23 AAU91166	Pantothenate kinas
12	232	18.3	455	20 AAY38617	Neisseria gonorrhoe
13	232	18.3	455	21 AAY74908	Neisseria gonorrhoe
14	232	18.3	460	23 AAU91167	Pantothenate kinas
15	232	18.3	592	20 AAY38618	Neisseria gonorrhoe
16	232	18.3	592	21 AAY74911	Neisseria gonorrhoe
17	210	16.5	265	23 AAU91151	Streptomyces coeli
18	192	15.1	189	20 AAY38614	Neisseria meningit
19	181	14.3	244	23 AAU91168	Pantothenate kinas
20	168.5	13.3	242	23 AAU91180	Pantothenate kinas
21	163	12.8	267	23 AAU91162	Bordetella pertussis
22	160.5	12.6	272	23 AAU91152	Mycobacterium tube
23	159	12.5	255	23 AAU91154	Geobacter sulfurre
24	157.5	12.4	272	22 AAU91150	Mycobacterium tube
25	156	12.3	250	23 AAU91150	Clostridium acetob
26	154	12.1	258	22 AAU01243	B. subtilis novel
27	154	12.1	258	23 AAU91149	Bacillus subtilis
28	153.5	12.1	260	23 AAU91173	Pantothenate kinas
29	144.5	11.4	223	19 AAU98422	H. pylori GHPO 344
30	144.5	11.4	223	23 AAU91161	Helicobacter pylor
31	144.5	11.4	223	23 AAU91181	Pantothenate kinas
32	142	11.2	262	23 AAU91170	Pantothenate kinas
33	139	11.0	258	23 AAU91172	Pantothenate kinas
34	134	10.6	219	23 AAU91176	Pantothenate kinas
35	131	10.3	256	23 AAU91175	Pantothenate kinas
36	130.5	10.3	257	23 AAU91174	Pantothenate kinas
37	129.5	10.2	233	23 AAU91163	Pantothenate kinas
38	124	9.8	258	23 AAU91153	Rhodobacter capsul
39	118	9.3	257	23 AAU91160	Synechocystis pant
40	118	9.3	273	23 AAU91157	Treponema pallidum
41	117.5	9.3	246	23 AAU91156	Thermotoga maritim
42	117	9.2	253	22 AAB79980	Corynebacterium g1
43	117	9.2	253	22 AAB80019	Corynebacterium g1
44	117	9.2	277	22 AAG90978	C. glutamicum prote
45	114.5	9.0	209	23 AAU91165	Pantothenate kinas

ALIGNMENTS

RESULT 1
AAU91178
ID AAU91178 standard; Protein; 249 AA.
XX
AC AAU91178;
XX
DT 05-JUN-2002. (first entry)
XX
DE Pantothenate kinase (Coax) #16.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Pseudomonas putida.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum ER, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54199.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein
 XX Claim 10; Page 108-109; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX Sequence 249 AA;
 SQ
 Query Match 100.0%; Score 1269; DB 23; Length 249;
 Best Local Similarity 100.0%; Pred. No. 3.3e-135;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MILEDCGNSFKRWIRHIVADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSRSSE 60
 1 MILEDCGNSFKRWIRHIVADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSRSSE 60
 QY 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 Db 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 QY 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 Db 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 QY 181 TVEAVERGCVLMQGFAYTQLEQARVLWGEEFTVLTGGDAPLVRAALPOARVVPDLVVF 240
 181 TVEAVERGCVLMQGFAYTQLEQARVLWGEEFTVLTGGDAPLVRAALPOARVVPDLVVF 240
 Db 241 GLAMACPLD 249
 241 GLAMACPLD 249
 QY 241 GLAMACPLD 249
 241 GLAMACPLD 249
 Db 241 GLAMACPLD 249
 241 GLAMACPLD 249
 RESULT 2
 AAU91182
 ID AAU91182 standard; Protein; 249 AA.
 XX AAU91182;
 AC AAU91182;
 DT 05-JUN-2002 (first entry)
 XX Pantothenate kinase (Coax) #20.
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX Pseudomonas syringae.
 OS WO200216601-A2.
 XX WO200216601-A2.
 PN 28-FEB-2002.
 XX 24-AUG-2001; 2001WO-US26531.
 XX 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.
 DR N-PSDB; ABK54203.
 XX

PT Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein
 XX Claim 10; Page 113-114; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX Sequence 249 AA;
 SQ
 Query Match 66.9%; Score 849; DB 23; Length 249;
 Best Local Similarity 64.9%; Pred. No. 1.3e-87;
 Matches 161; Conservative 32; Mismatches 55; Indels 0; Gaps 0;
 QY 1 MILEDCGNSFKRWIRHIVADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSRSSE 60
 1 MILEDCGNSFKRWIRHIVADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSRSSE 60
 Db 1 MILEDCGNSFKRWIRHIVADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSRSSE 60
 1 MILEDCGNSFKRWIRHIVADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSRSSE 60
 QY 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 Db 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 61 ETDALCALIAQAFVAQVAKVHPVREMGVNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120
 QY 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 Db 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 121 IDLGTAAKADFSADGHEHGGYICPGMLRMSOLRTHRRIRYDDASAEALSSLSPPGRS 180
 QY 181 TVEAVERGCVLMQGFAYTQLEQARVLWGEEFTVLTGGDAPLVRAALPOARVVPDLVVF 240
 181 TVEAVERGCVLMQGFAYTQLEQARVLWGEEFTVLTGGDAPLVRAALPOARVVPDLVVF 240
 Db 241 GLAMACPLD 248
 241 GLAMACPLD 248
 QY 241 GLAMACPLD 248
 241 GLAMACPLD 248
 Db 241 GLAMACPLD 248
 241 GLAMACPLD 248
 RESULT 3
 AAU91164
 ID AAU91164 standard; Protein; 248 AA.
 XX AAU91164;
 AC AAU91164;
 DT 05-JUN-2002 (first entry)
 XX Pantothenate kinase (Coax) #2.
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX Pseudomonas aeruginosa.
 OS WO200216601-A2.
 XX WO200216601-A2.
 PN 28-FEB-2002.
 XX 24-AUG-2001; 2001WO-US26531.
 XX 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.
 DR N-PSDB; ABK54203.

XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
XX Claim 6; Page 83-84; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 248 AA;
SQ
Query Match 62.6%; Score 794.5; DB 23; Length 248;
Best Local Similarity 63.5%; Pred. No. 1.9e-81;
Matches 158; Conservative 31; Mismatches 59; Indels 1; Gaps 1;
QY 1 MILELDCGNSFKRWVHVADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSVRSEE 60
DB 1 MILELDCGNSLTKRWVIEGARSV-AGGLAESSDALVEQLTSQALPVRACRLSVSRSEQ 59
QY 61 ETDALCALIAQAFVAQVAHVPRVEMAGVRNGYDDYQRLGMDRWLAALGAPHLAKACLV 120
DB 60 ETSQVLALEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLVAALHAKKACLV 119
QY 121 IDLGTAAKADFSADGEHLGGYICPGMPLMRSQLRTHRRIRYDDASAEALSSLSPGRS 180
DB 120 IDLGTAVTSDLVADGVHLGGYICPGMTLMRSQLRTHRRIRYDDAEARRALASLPQQA 179
QY 181 TVEAVERGCVLMQGFATVQLEQARVLMGEEFTVLTGGDAPLVRAALPQARVVPDLVFFV 240
DB 180 TAEAEVERGCLLMRGFVEQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDVFFV 239
QY 241 GLAMACPLD 249
DB 240 GLALACPTE 248
RESULT 4
AAV38615
D AAV38615 standard; Protein; 592 AA.
XX
XX AAV38615;
AC
XX
XX 08-OCT-1999 (first entry)
XX
XX Neisseria meningitidis antigen encoded by ORF61.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
XX Neisseria meningitidis.
XX
XX WO9924578-A2.
PN
XX
XX 20-MAY-1999.
PD
XX
XX 09-OCT-1998; 98WB-IB01665.
PF
XX
XX 01-SEP-1998; 98GB-0019016.
PR
XX 06-NOV-1997; 97GB-0023516.
PR
XX 14-NOV-1997; 97GB-0024190.
PR
XX 18-NOV-1997; 97GB-0024386.
PR
XX 27-NOV-1997; 97GB-0025158.
PR
XX 10-DEC-1997; 97GB-0026147.
PR
XX 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.
PA Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;
PI
XX WPI; 1999-327407/27.
DR N-PSDB; AA212073.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 172; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
XX Sequence 592 AA;
SQ
Query Match 19.3%; Score 245; DB 20; Length 592;
Best Local Similarity 30.1%; Pred. No. 1.2e-18;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
QY 3 LEELDCGNSFKRWVHVADAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
DB 341 LEELDCGNSLTKW-----AWVNGTTFATVGSAPYRDLSPGLGAWEAKDGNVRIYGC-- 391
QY 54 VSVRSEETDLCALIAQAFVAQVAHVPR-----EMAGVRNGYDDYQRLGMDRWLA 107
DB 392 -----AVCGEFKA-QVQELARKIEWLPSSAQALGINNHRRHFEHSGDRWFA 440
QY 108 LGAFHLAKGACLVLDLGTAAKADFSADGEHLGGYICPGMPLMRSQLRTHRRIRI----- 161
DB 441 LGSRRFSRNACVWVSCGTAVTVDALTDGHLGGTTPGPHLMKESLAVRTANLNRHAG 500
QY 162 RYD-DASAEALSSLSPGRSTVEAVERGCVLMQGFATVQLEQARVLMGEEFTVLTGGD 220
DB 501 RYFPFTTTGNVAS-----GMMDAV-CGSVMMHG----RLKE-KTGAGKPDVVIITGG 549
QY 221 APLVRAALPQA-----RVVPDLVFFVL 242
DB 550 AAKVAELPPAFIAENTVRVADNLVIYGL 578
RESULT 5
AAV74912
ID AAV74912 standard; Protein; 592 AA.
XX
XX AAV74912;
AC
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
OS
XX WO9957280-A2.
PN
XX
XX 11-NOV-1999.
PD
XX
XX 30-APR-1999; 99WO-US09346.
PF
XX
XX 01-MAY-1998; 98US-0083758.
PR
XX 31-JUL-1998; 98US-0094869.
PR

PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 98US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR N-PSDB; AA253674.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 Claim 2; Page 714; 1453pp; English.

AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
 represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 PCR primers used in the exemplification of the present invention. The
 polypeptides, the polynucleotides, antibodies and compositions of
 the invention can be used as vaccines, as diagnostic reagents, and as
 immunogenic compositions. The polypeptides can be used in the
 manufacture of medicaments for treating or preventing infection due to
Neisseria meningitidis (e.g. meningitis and septicemia), to detect the
 presence of *Neisseria meningitidis* bacteria, or to raise antibodies. They may also
 be used to screen for agonists or antagonists, which may themselves
 have use as antibacterial agents. The polynucleotides of the invention
 may also be used in gene therapy protocols.

XX Sequence 592 AA;

Query Match 19.3%; Score 245; DB 21; Length 592;
 Best Local Similarity 30.1%; Pred. No. 1.2e-18;
 Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
 QY 3 LEIDGNSPIKRVHIVADAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
 DB 341 LLLDGGNSRLKW-----AWENGTFATVGSAPYRDLSPLGAWEAKADGNVRIYGC-- 391
 QY 54 VSVRSEETDLCALIAQAFVAQVAVHPVR-----EMAGVRNGYDDYQRLGMDRLAA 107
 DB 392 -----AVCGEFKKA-QVQEQRLARKIEWLPSSAQALGIRNHYRHPHEHGSDFWNA 440
 108 LGAFHLAKACLVLDLGTAAKADFSADGEHLGGYICPGMPLMRSOLRTHTRRI----- 161
 DB 441 LGSRRFSRNACVSVSGTAVTDALDDGHYLGTTIMPGFHLMKESLAVRTANLNHACK 500
 QY 162 RYD-DASABRALSSLSGPRSTVEAVERGCVLMQGFAYTQLEQARVLMGEETFTVTGSD 220
 DB 501 RYFPFTTTCNAVAS-----GMDAV-CGSVMAMHG-----RLKE-KTGACKPVDVITGGG 549
 QY 221 APLVRAALPQA-----RVVPDLVFLVGL 242
 DB 550 AAKVAELPPAFLAENTVRVADNLVIYGL 578

RESULT 6
 AAU91169
 ID AAU91169 standard; Protein; 592 AA.
 XX
 AC AAU91169;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Pantothenate kinase (Coax) #7.
 XX

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN WO200216601-A2.
 XX
 XX 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26531.
 XX
 PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX
 PI Yocum RR, Patterson TA;
 XX
 DR WPI: 2002-269358/31.
 DR N-PSDB; ABK54190.
 XX
 PT Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 PS Claim 6; Page 96-98; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.

XX Sequence 592 AA;

Query Match 19.3%; Score 245; DB 23; Length 592;
 Best Local Similarity 30.1%; Pred. No. 1.2e-18;
 Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
 QY 3 LEIDGNSPIKRVHIVADAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
 DB 341 LLLDGGNSRLKW-----AWENGTFATVGSAPYRDLSPLGAWEAKADGNVRIYGC-- 391
 QY 54 VSVRSEETDLCALIAQAFVAQVAVHPVR-----EMAGVRNGYDDYQRLGMDRLAA 107
 DB 392 -----AVCGEFKKA-QVQEQRLARKIEWLPSSAQALGIRNHYRHPHEHGSDFWNA 440
 QY 108 LGAFHLAKACLVLDLGTAAKADFSADGEHLGGYICPGMPLMRSOLRTHTRRI----- 161
 DB 441 LGSRRFSRNACVSVSGTAVTDALDDGHYLGTTIMPGFHLMKESLAVRTANLNHACK 500
 QY 162 RYD-DASABRALSSLSGPRSTVEAVERGCVLMQGFAYTQLEQARVLMGEETFTVTGSD 220
 DB 501 RYFPFTTTCNAVAS-----GMDAV-CGSVMAMHG-----RLKE-KTGACKPVDVITGGG 549
 QY 221 APLVRAALPQA-----RVVPDLVFLVGL 242
 DB 550 AAKVAELPPAFLAENTVRVADNLVIYGL 578

RESULT 7
 AAU74909
 ID AAU74909 standard; Protein; 389 AA.
 XX
 AC AAU74909;
 XX
 DT 21-MAR-2000 (first entry)
 XX

DE Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1292.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX (CHIRON) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
DR N-PSDB; AAZ53671.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 2; Page 710; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 389 AA;
XX
XX Query Match 19.3%; Score 244.5; DB 21; Length 389;
XX Best Local Similarity 30.0%; Pred. No. 7.7e-19;
XX Matches 81; Conservative 40; Mismatches 88; Indels 61; Gaps 12;
XX
XX 3 LEIDCGNSFKRWIRHVADAVIEGGGIVDSQALVAEVAAL-----ASVRLTGCRI 53
XX 138 LLLDGGNSRLKW-----AWVNGTATVGSAPYRDLSPIGAWEAEKADGNVRIVGC--'188
XX
XX 54 VSVRSEETDICALIAQAFVQAKVAHPV-----REMAVGNGYDDYQRLGMDRWLA 106
XX 189 -----AVCGEFKKA-QVQEQLARKIEWLPSSAQALFGIKNHRYRHPPEEHSRWFN 237
XX
XX 107 ALGAFHLAKGACLVLDLGTAAKADFSVADGHEHGGYICPGMPLMRSLRTHTRRI-----161
XX 238 ALGSRFRSRNACVSVVCGTATVTDALDDGHYLGTTMPGFHLMKESLAVTANLNHAG 297
XX
XX 162 -RYD-DASAEARLSSLSFGSRSTVEAVERGCVLMQLQGFAYTQLEQARVLMGEEFTVITGG 219
XX 298 KRYFPFTTGNVAVS-----GMMDAV-CGSVMMHG-----RLKE-KTGACKPVDVITGG 346

QY 220 DAPLVRALPOA-----RVVPDLVFVGL 242
DB 347 GAAKVAELPAPFLAENTVRVADNLVIYGL 376

RESULT 8

AAZ74910
ID AAY74910 standard; Protein; 455 AA.

XX AAY74910;

XX 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.

XX Neisseria meningitidis.

XX WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 25-FEB-1999; 99US-0121528.

XX (CHIRON) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; AAZ53672.

XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics

XX Claim 2; Page 711; 1453pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 19.1%; Score 243; DB 21; Length 455;
Best Local Similarity 30.1%; Pred. No. 1.4e-18;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;

QY 3 LEIDCGNSFKRWIRHVADAVIEGGGIVDSQALVAEVAAL-----ASVRLTGCRI 53

```

Db 204 LLLDGNLSRLK-----AWYENGTFATVGSAPYRDLSPLAGAEWAQKVDGNVRIVGC-- 254
QY 54 VSVRSEETDLCALIAQAFAYQAKVAHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107
Db 255 -----AVCGEFKKA-QVQEQIARKIEWLPSSAQALGIRNHYRHPHEHGSDFWFA 303
QY 108 LGAFHLAKGACLVLDLGTAAKADFSADGEHLGGYICPGMPLMRSQLRTHTRRI----- 161
Db 304 LGSRRFSRNACVVGSGTAVTVDALTDGHLGGTIMPGFHLMKESLAVRTANLNRHAGK 363
QY 162 RYD-DASAEALSSLSPGSRSTVEAVERGCVLMQGFAYTQLEQARVLMGEEETVFLTGGD 220
Db 364 RYPPFTTTGNVAVS-----GMMDAV-CGSVMMHG-----RLKE-KTGAGKPDVDVIITGGG 412
QY 221 APLVRAALPOA-----RVVDPDLVFEVL 242
Db 413 AAKVAELPPAFLAENTVRVADNLVIHGL 441

RESULT 9
AY38616
AA38616 standard; Protein; 592 AA.
XX AC AY38616;
DT DT 08-OCT-1999 (first entry)
XX DE Neisseria meningitidis strain A antigen encoded by ORF61.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX OS Neisseria meningitidis.
XX PN WO924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1998; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
XX Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
XX DR N-PSDB; AA212074.
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX Claim 4; Page 173; 524pp; English.
XX Amino acid sequences AY38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX Sequence 592 AA;
XX Query Match 19.1%; Score 243; DB 20; Length 592;

Best Local Similarity 30.1%; Pred. No. 2e-18;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
QY 3 LELDCGNSFIKWRVTHVADAVIEGGIVDSOALVAEVAAL-----ASVRLTGCR 53
Db 341 LLLDGNLSRLK-----AWYENGTFATVGSAPYRDLSPLAGAEWAQKVDGNVRIVGC-- 391
QY 54 VSVRSEETDLCALIAQAFAYQAKVAHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107
Db 392 -----AVCGEFKKA-QVQEQIARKIEWLPSSAQALGIRNHYRHPHEHGSDFWFA 440
QY 108 LGAFHLAKGACLVLDLGTAAKADFSADGEHLGGYICPGMPLMRSQLRTHTRRI----- 161
Db 441 LGSRRFSRNACVVGSGTAVTVDALTDGHLGGTIMPGFHLMKESLAVRTANLNRHAGK 500
QY 162 RYD-DASAEALSSLSPGSRSTVEAVERGCVLMQGFAYTQLEQARVLMGEEETVFLTGGD 220
Db 501 RYPPFTTTGNVAVS-----GMMDAV-CGSVMMHG-----RLKE-KTGAGKPDVDVIITGGG 549
QY 221 APLVRAALPOA-----RVVDPDLVFEVL 242
Db 550 AAKVAELPPAFLAENTVRVADNLVIHGL 578

RESULT 10
AY74913
AA74913 standard; Protein; 592 AA.
XX AC AY74913;
DT DT 21-MAR-2000 (first entry)
XX DE Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1300.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy.
XX OS Neisseria meningitidis.
XX PN WO9957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX (CHIR-) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
XX PI Petersen J, Pizza M, Rappuoli R, Rattl G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX DR N-PSDB; AA253675.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX Claim 2; Page 715; 1453pp; English.
XX AA253015 to AA254536, AA254577 to AA254615, and AA254537 to AA254576
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254573 represent

```


PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26531.
 XX
 PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX
 PI Yocum RR, Patterson TA;
 XX
 XX WPI; 2002-269358/31.
 DR N-PSDB; ABK54188.
 DR
 XX
 PT Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 XX Claim 8; Page 92-94; 128pp; English.
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 460 AA;
 Query Match 18.3%; Score 232; DB 23; Length 460;
 Best Local Similarity 29.0%; Pred. No. 2.5e-17;
 Matches 78; Conservative 41; Mismatches 90; Indels 60; Gaps 12;
 QY 3 LELDCGNSFIKRWIYHADVAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
 DB 209 LLEGGNSRLKW-----AWVNGTFATVGSAPYRDLSPGLAEWAERKADGNVRIVGC-- 259
 QY 54 VSVRSEETDLCALIAQAFVAVQAKVHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107
 DB 260 -----AVCGESKKA-QVKEQLARKIEWLPSAQAALGIRNHYRHPHEHSGDRWFNA 308
 QY 108 LGAFHLAKGACLVLDLGTAAKADFSADGHELGVCYICPGMPLMRSLRTHTRRI----- 161
 309 LGSRRFSRNACVVVSCGTAVTVTDALTDGHLGGTIMPGFHLMKESLAVRTANLNRPAGK 368
 QY 162 RYD-DASAERALSLSLSPGRSTVEAVERGCVLMQLQGFAYTQLEQARVLWGEFTVFLTGGD 220
 DB 369 RYPFPTTTGNAS-----GMMDAV-CGSIIMMHG-----RLKEKNGA-GKPDVVIITGGG 417
 QY 221 APLVRAALPQA-----RVVPDLVYFVGL 242
 DB 418 AAKVAELPPAPLAENTVRVADNLVHGL 446
 RESULT 15
 ID AAY38618 standard; Protein; 592 AA.
 XX
 AC AAY38618;
 XX
 DT 08-OCT-1999 (first entry)
 XX
 XX Neisseria gonorrhoeae antigenic protein encoded by ORF61.
 XX
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
 XX
 OS Neisseria gonorrhoeae.
 XX

PN W09924578-A2.
 XX
 PD 20-MAY-1999.
 XX
 PF 09-OCT-1998; 98WO-IB01665.
 XX
 PR 01-SEP-1998; 98GB-0019016.
 PR 06-NOV-1997; 97GB-0023516.
 PR 14-NOV-1997; 97GB-0024190.
 PR 18-NOV-1997; 97GB-0024386.
 PR 27-NOV-1997; 97GB-0025158.
 PR 10-DEC-1997; 97GB-0026147.
 PR 14-JAN-1998; 98GB-0000759.
 XX
 XX (CHIR-) CHIRON SPA.
 PA
 XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
 PI
 DR WPI; 1999-327407/27.
 DR N-PSDB; AAZ12075.
 DR
 XX
 PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
 PT diagnosis, treatment and prevention of infection
 XX
 PS Claim 4; Page 175; 524pp; English.
 CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
 CC and N. gonorrhoeae antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AAZ11972-212358. The antigenic proteins, or
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of Neisseria
 CC infections, such as meningitis, septicaemia and gonorrhea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.
 XX
 SQ Sequence 592 AA;
 Query Match 18.3%; Score 232; DB 20; Length 592;
 Best Local Similarity 29.0%; Pred. No. 3.6e-17;
 Matches 78; Conservative 41; Mismatches 90; Indels 60; Gaps 12;
 QY 3 LELDCGNSFIKRWIYHADVAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
 DB 341 LLEGGNSRLKW-----AWVNGTFATVGSAPYRDLSPGLAEWAERKADGNVRIVGC-- 391
 QY 54 VSVRSEETDLCALIAQAFVAVQAKVHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107
 DB 392 -----AVCGESKKA-QVKEQLARKIEWLPSAQAALGIRNHYRHPHEHSGDRWFNA 440
 QY 108 LGAFHLAKGACLVLDLGTAAKADFSADGHELGVCYICPGMPLMRSLRTHTRRI----- 161
 441 LGSRRFSRNACVVVSCGTAVTVTDALTDGHLGGTIMPGFHLMKESLAVRTANLNRPAGK 500
 QY 162 RYD-DASAERALSLSLSPGRSTVEAVERGCVLMQLQGFAYTQLEQARVLWGEFTVFLTGGD 220
 DB 501 RYPFPTTTGNAS-----GMMDAV-CGSIIMMHG-----RLKEKNGA-GKPDVVIITGGG 549
 QY 221 APLVRAALPQA-----RVVPDLVYFVGL 242
 DB 550 AAKVAELPPAPLAENTVRVADNLVHGL 578
 Search completed: June 24, 2003, 22:08:40
 Job time : 29.1855 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.3982 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-3
Perfect score: 1257
Sequence: 1 NKRAAFMLLLFLRSVLKVL.....GRTRTSLVATGGLAKLIN 250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : PublishedApplications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	1257	100.0	250	9	US-09-813-453A-3
2	695	55.3	256	9	US-09-813-453A-55
3	665.5	52.9	262	9	US-09-813-453A-45
4	648.5	51.6	254	9	US-09-813-453A-47
5	646.5	51.4	258	9	US-09-813-453A-2
6	631.5	50.2	255	9	US-09-813-453A-7
7	629.5	50.1	258	9	US-09-813-453A-49
8	623	49.6	233	9	US-09-813-453A-17
9	513	40.8	219	9	US-09-813-453A-57
10	498.5	39.7	265	9	US-09-813-453A-4
11	460.5	36.6	260	9	US-09-813-453A-51
12	428.5	34.1	258	9	US-09-813-453A-6
13	392	31.2	272	9	US-09-712-363-276
14	389	30.9	272	9	US-09-813-453A-5
15	337	26.8	262	9	US-09-813-453A-8
16	333.5	26.5	246	9	US-09-813-453A-9
17	327.5	26.1	212	9	US-09-813-453A-59
18	282.5	22.5	273	9	US-09-813-453A-10
19	262	20.8	262	9	US-09-813-453A-11

20	229.5	18.3	257	9	US-09-813-453A-53
21	169	13.4	241	9	US-09-813-453A-63
22	169	13.4	460	9	US-09-813-453A-39
23	168	13.4	592	9	US-09-813-453A-22
24	168	13.4	592	9	US-09-813-453A-43
25	167.5	13.3	244	9	US-09-813-453A-41
26	159.5	12.7	248	9	US-09-813-453A-20
27	156	12.4	249	9	US-09-813-453A-61
28	141	11.2	267	9	US-09-813-453A-15
29	136.5	10.9	249	9	US-09-813-453A-70
30	123.5	9.8	229	9	US-09-813-453A-12
31	121.5	9.7	257	9	US-09-813-453A-13
32	108.5	8.6	242	9	US-09-813-453A-65
33	102.5	8.2	223	9	US-09-895-913A-74
34	102.5	8.2	223	9	US-09-813-453A-14
35	102.5	8.2	223	9	US-09-813-453A-67
36	86	6.8	872	9	US-10-047-260-38
37	85.5	6.8	542	10	US-09-741-669-414
38	84.5	6.7	542	9	US-09-308-207-16
39	83.5	6.6	856	10	US-09-815-242-11489
40	82	6.5	396	9	US-10-285-074-2
41	82	6.5	396	9	US-10-285-074-4
42	82	6.5	856	10	US-09-815-242-11310
43	81	6.4	533	9	US-10-108-605-193
44	79.5	6.3	607	9	US-10-072-436-11
45	79	6.3	854	10	US-09-815-242-5192

ALIGNMENTS

RESULT 1

US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: CGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

Query Match	100.0%	Score 1257	DB 9	Length 250
Best Local Similarity	100.0%	Pred. No. 4.9e-120		
Matches 250	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	NKRAAFMLLLFLRSVLKVLVDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQV	60	
Db	1	NKRAAFMLLLFLRSVLKVLVDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQV	60	
Qy	61	MNLFQDDKLPDTVEGVIISVVPNIMYSLEHMKRYEKNPLVVGPGIKTGINKYDNP	120	
Db	61	MNLFQDDKLPDTVEGVIISVVPNIMYSLEHMKRYEKNPLVVGPGIKTGINKYDNP	120	
Qy	121	KEVGADRVNVAHAHEIYKRSLLIIDFGTATTCFVRENGDYLGGATCPGKVSSEALFE	180	
Db	121	KEVGADRVNVAHAHEIYKRSLLIIDFGTATTCFVRENGDYLGGATCPGKVSSEALFE	180	
Qy	161	KAALPRVELIKPAYATCKNTYISSIQSGIVYVLRQVYLFKEKLENLPGRRRTTSVLV	240	
Db	161	KAALPRVELIKPAYATCKNTYISSIQSGIVYVLRQVYLFKEKLENLPGRRRTTSVLV	240	

Db 181 KAAKLPVELIKPAYAICKNTISSIOSGIVYRLROVKYLFKELKENLPDGRTRTSLVL 240
QY 241 ATGGLAKLIN 250
1111111111
Db 241 ATGGLAKLIN 250
1111111111
RESULT 2
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRF
; ORGANISM: Clostridium difficile
US-09-813-453A-55

Query Match 55.3%; Score 695; DB 9; Length 256;
Best Local Similarity 59.0%; Pred. No. 9.2e-63;
Matches 138; Conservative 37; Mismatches 51; Indels 8; Gaps 2;
QY 18 VILVDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQVMNLFQDQKLDPTLVEGV 77
Db 1 MLVFDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQVMNLFQDQKLDPTLVEGV 60
QY 78 IISVVVPNTMYSLEHMIRKYFKINPLVVGPGIKTGNIKYNKPEKVGADRVNVAHAHEI 137
Db 61 IISVVVPNTMYSLEHMIRKYFKINPLVVGPGIKTGNIKYNKPEKVGADRVNVAHAHEI 120
QY 138 YKRSIIIDFGTATTCFAVRENGDYLGGAIKPCGKIVSSSEALFEKAAKLPVELIKPAYAI 197
Db 121 YGAPSLVDGTTATTCFAVRENGDYLGGAIKPCGKIVSSSEALFEKAAKLPVELIKPAYAI 180
QY 198 CKNTISSIOSGIVYRLROVKYLFKELKE--NLPDGRTRTSLVATGLAKLI 249
Db 181 CKSTVSAMSGIYGYVGLVDKIIISIMKKELNCDDVK-----VIATGLAKLI 228

RESULT 3
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRF

; ORGANISM: Bacillus anthracis
US-09-813-453A-45
Query Match 52.9%; Score 665.5; DB 9; Length 262;
Best Local Similarity 54.9%; Pred. No. 9.6e-60;
Matches 128; Conservative 45; Mismatches 55; Indels 5; Gaps 1;
QY 18 VILVDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQVMNLFQDQKLDPTLVEGV 77
Db 1 MIFVLDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQVMNLFQDQKLDPTLVEGV 60
QY 78 IISVVVPNTMYSLEHMIRKYFKINPLVVGPGIKTGNIKYNKPEKVGADRVNVAHAHEI 137
Db 61 IISVVVPNTMYSLEHMIRKYFKINPLVVGPGIKTGNIKYNKPEKVGADRVNVAHAHEI 120
QY 138 YKRSIIIDFGTATTCFAVRENGDYLGGAIKPCGKIVSSSEALFEKAAKLPVELIKPAYAI 197
Db 121 YGAPSLVDGTTATTCFAVRENGDYLGGAIKPCGKIVSSSEALFEKAAKLPVELIKPAYAI 180
QY 198 CKNTISSIOSGIVYRLROVKYLFKELKENLPDGRTRTSLVATGLAKLI 250
Db 181 GKNVTSAMSGIYGYVGVGGIVGIVKRMKEAKQEPK-----VIATGLAKLI 228

RESULT 4
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRF
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 51.6%; Score 648.5; DB 9; Length 254;
Best Local Similarity 53.6%; Pred. No. 5e-58;
Matches 126; Conservative 41; Mismatches 57; Indels 11; Gaps 2;
QY 18 VILVDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQVMNLFQDQKLDPTLVEGV 77
Db 1 MILVIDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQVMNLFQDQKLDPTLVEGV 60
QY 78 IISVVVPNTMYSLEHMIRKYFKINPLVVGPGIKTGNIKYNKPEKVGADRVNVAHAHEI 137
Db 61 VISSVPPNMFSLQCKKCYFHVTPMIGGKIGKTNKYNKPEKVGADRVNVAHAHEI 120
QY 138 YKRSIIIDFGTATTCFAVRENGDYLGGAIKPCGKIVSSSEALFEKAAKLPVELIKPAYAI 197
Db 121 YGAPSLVDGTTATTCFAVRENGDYLGGAIKPCGKIVSSSEALFEKAAKLPVELIKPAYAI 180
QY 198 CKNTISSIOSGIVYRLROVKYLFKELK---ENLPDGRTRTSLVATGLAKLI 249
Db 181 GTNTIDSMQSGIFGYGVQDGVGVKRMKAQAESEPK-----VIATGLAKLI 227

RESULT 5
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1


```

; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-813-453A-2

Query Match      51.4%; Score 646.5; DB 9; Length 258;
Best Local Similarity 53.9%; Pred. No. 8.1e-58;
Matches 125; Conservative 46; Mismatches 56; Indels 5; Gaps 1;

Qy 18 VILVDGNTNVLGIYNDTKLTAEWRSLTDVLSADEYGIQVNMFLQDLDPTLVBEV 77
Db 1 LLLVIDGNTNVLGVYDGHGLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60

Qy 78 IISVVPNIMYSLEHMIRKYPKINPLVVGPGIKTGKINIKYDNPKVEGADRIVNAVAHEI 137
Db 61 IISVVPNIMYSLEHMIRKYPKINPLVVGPGIKTGKINIKYDNPKVEGADRIVNAVAHEI 120

Qy 138 YKRSIIIDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 197
Db 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPIGIMISTEALFARAALPRIEIRPDNII 180

Qy 198 KNTISSQSGIVRYLRQVKYLFEPKLENLPDGRTRTSLVATGGLAKLI 249
Db 181 GKNTVSAMQSGILGYGVQVEGIVKRMKQAKQDLK-----VIATGGLAPLI 227

```

```

RESULT 6
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
; US-09-813-453A-7

Query Match      50.2%; Score 631.5; DB 9; Length 255;
Best Local Similarity 54.7%; Pred. No. 2.7e-56;
Matches 127; Conservative 41; Mismatches 59; Indels 5; Gaps 1;

Qy 18 VILVDGNTNVLGIYNDTKLTAEWRSLTDVLSADEYGIQVNMFLQDLDPTLVBEV 77
Db 1 MLLVIDGNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60

```

```

Qy 78 IISVVPNIMYSLEHMIRKYPKINPLVVGPGIKTGKINIKYDNPKVEGADRIVNAVAHEI 137
Db 61 IISVVPNIMYSLEHMIRKYPKINPLVVGPGIKTGKINIKYDNPKVEGADRIVNAVAHEI 120

Qy 138 YKRSIIIDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 197
Db 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPIGIMISTEALFARAALPRIEIRPDNII 180

Qy 198 KNTISSQSGIVRYLRQVKYLFEPKLENLPDGRTRTSLVATGGLAKLI 249
Db 181 GKNTVSAMQSGILGYGVQVEGIVKRMKQAKQDLK-----VIATGGLAPLI 227

RESULT 7
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
; US-09-813-453A-49

```

```

Query Match      50.1%; Score 629.5; DB 9; Length 258;
Best Local Similarity 53.8%; Pred. No. 4.4e-56;
Matches 126; Conservative 40; Mismatches 59; Indels 9; Gaps 2;

Qy 18 VILVDGNTNVLGIYNDTKLTAEWRSLTDVLSADEYGIQVNMFLQDLDPTLVBEV 77
Db 1 MIFVLDGNTNVLGVYDGHGLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60

Qy 78 IISVVPNIMYSLEHMIRKYPKINPLVVGPGIKTGKINIKYDNPKVEGADRIVNAVAHEI 137
Db 61 IISVVPNIMYSLEHMIRKYPKINPLVVGPGIKTGKINIKYDNPKVEGADRIVNAVAHEI 120

Qy 138 YKRSIIIDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 197
Db 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPIGIMISTEALFARAALPRIEIRPDNII 180

Qy 198 KNTISSQSGIVRYLRQVKYLFEPKLENLPDGRTRTSLVATGGLAKLI 249
Db 181 GKNTVSAMQSGILGYGVQVEGIVKRMKQAKQDLK-----PKVIATGGLAPLI 227

RESULT 8
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569

```


Matches 92; Conservative 53; Mismatches 81; Indels 7; Gaps 2;
Qy 18 VILVDVGNINVLGIYNDKLTAEWRLSTDVLSADEYGIQVNNLFQDDKLDPTLVEGV 77
Db 1 MLLAEQGNNTWTFPAIHGASWAOWRSATSTRTADEYVWLSQLLSMGLGFRAIDAV 60
Qy 78 IISVVPNIMYSLEHMTIRKYFKINPLVVGPIKGTINIKYDNPKVEGADRIYNAVAHEI 137
Db 61 IISVVPQSFNLRNLSRRYFNVEPLVIGENAKLIDVRIEKPSEACADRLVAIGAAMV 120
Qy 138 YKRSLLIIDGTATTCFAVRENGDYLGGACPGIKVSSEALFEKAALPRVELIKPA--Y 195
Db 121 YPGLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180
Qy 196 AICKNTISSQSGIVYRLQVKYLFKELKENLPDGRTRTSLVLTATGGLAKL 248
Db 181 IVGTDVTSAMQSGVFWGYSLIEGLVARIK-----AERGPMVIATGGVASL 228

RESULT 12

US-09-813-453A-6
; Sequence 6, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

Query Match 34.1%; Score 428.5; DB 9; Length 258;
Best Local Similarity 39.7%; Pred. No. 1.3e-35;
Matches 93; Conservative 45; Mismatches 89; Indels 7; Gaps 3;

Qy 18 VILVDVGNINVLGIYNDKLTAEWRLSTDVLSADEYGIQVNNLFQDDKLDPTLVEGV 77
Db 1 MLLCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLTMLQKLGQRISE-A 59
Qy 78 IISVVPNIMYSLEHMTIRKYFKINPLVVG-PGIKGTINIKYDNPKVEGADRIYNAVAHE 136
Db 60 IISTAPRVNLRVLCNRFDCRPYVVGPGGELPVPAPDGTTPDRLVNTVAGID 119
Qy 137 YKRSLLIIDGTATTCFAVRENGDYLGGACPGIKVSSEALFEKAALPRVELIKPAYA 196
Db 120 RHGGDLVDFGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 179
Qy 197 ICKNTISSQSGIVYRLQVKYLFKELKENLPDGRTRTSLVLTATGGLAKLIN 250
Db 180 IGTNTVACIOSGVWYIGLVEGIVRQIRM-----ERDRPMKVIATGGLASLFD 228

RESULT 13

US-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND

; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-276

Query Match 31.2%; Score 392; DB 9; Length 272;
Best Local Similarity 34.6%; Pred. No. 7.4e-32;
Matches 82; Conservative 62; Mismatches 85; Indels 8; Gaps 3;

Qy 18 VILVDVGNINVLGIYNN----DTKLTAEWRLSTDVLSADEYGIQVNNLFQDDKLDPTL 73
Db 1 VLLAIDVRNTHTVVGLLSGMKEHAKVQQMRIRTESEVTADALTIDGLIGEDS----ER 57
Qy 74 VEGVLISSVVPNIMYSLEHMTIRKYFKINP-LVVGPGIKGTINIKYDNPKVEGADRIYNAV 132
Db 58 LTGTALSTVPSVLHEVRIMLDQWPSVPHVIEPGVTGIPLVDPNPKVEGADRIYNAVCL 117
Qy 133 AAHEIYKRSLLIIDGTATTCFAVRENGDYLGGACPGIKVSSEALFEKAALPRVELIK 192
Db 118 AAVDRFRKAAIVVDFGSSICVDVVSAGFELGGAIFGVQVSSDAAAARSAAARVELAR 177
Qy 193 PAYAICKNTISSQSGIVYRLQVKYLFKELKENLPDGRTRTSLVLTATGGLAKLI 249
Db 178 PRSVYVGKNTVECMQAGAVFGFAGLVGLVGRIREDSVGFSDHVDVAIVATGHTAPLL 234

RESULT 14

US-09-813-453A-5
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis

US-09-813-453A-5

Query Match 30.9%; Score 389; DB 9; Length 272;
Best Local Similarity 34.2%; Pred. No. 1.5e-31;
Matches 81; Conservative 63; Mismatches 85; Indels 8; Gaps 3;
QY 18 VILVDVGNINVLGIYN---DTKLTAEWRLSTDVLSADEYGIQVMNLFQODKLDPTL 73
DB 1 MLLAIDVRNTHTVGLLSGKHAQVQWRIETSEVTADLALTDGLIGDS---ER 57
QY 74 VEGVISSVVPNTIMYSLHMKRYFKINP-LVVGPGIKTGINKYDNPKEVGADRIVNAV 132
DB 58 LTGTAALSTVPSLVHEVRIMLDQWSPVPHVLEPGVGTGIPLLVDNPKREVAGDRIVNC 117
QY 133 AAHEIYKRSLLIIDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALKPRVELIK 192
DB 118 AAYDRFKAIIVDFGSSICVDVWSAKGEFLGGAIAFGVQVSSDAAAARSALRRVELAR 177
QY 193 PAYAICKNTISSIQSGIVRYLRQVKYLFKELKENLPDGRRTTSLVLTATGGLAKLI 249
DB 178 PRSWGKNTVECMQAGAVEGAGLVGDLVGRIREVDVSGFSDVDHDAIVATGHTAPLL 234

RESULT 15

US-09-813-453A-8
; Sequence 8, Application US/09813453A
; Patent No. US2002016861A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 8
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Deinococcus radiopugnans
US-09-813-453A-8

Query Match 26.8%; Score 337; DB 9; Length 262;
Best Local Similarity 33.3%; Pred. No. 2.8e-26;
Matches 80; Conservative 49; Mismatches 89; Indels 22; Gaps 6;
QY 19 ILVLDVGNINVLGIYNDT-KLTAEWRLSTDVLSADEYGIQVMNLFQODKLDPTLV--- 74
DB 6 LLAVDIGNTTVLGLADASGALHTWRIETNREMLPDDLALQLHGLF-----TLGAP 58
QY 75 --EGVISSVVPNTIMYSLHMKRYFKINP-LVVGPGIKTGINKYDNPKEVGADRIVNAV 132
DB 59 IPRAAVLSSVAPPVGENYALALKRHPMIDAFVSAENLPDVTVELDTPGSGVADRLCNLF 118
QY 133 AAHEIYKRSLLIIDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALKPRVE 189
DB 119 GA-EKYLGLDYAVVVDFTSTNFDVVRGRFLGILATGAQVSADALFARAALKPRIT 177
QY 190 LTKPAYACKNTISSIQSGIVRYLRQVKYLFKELKENLPDGRRTTSLVLTATGGLAKLI 249
DB 178 LQAPETAICKNTVHALQSLVGLVGYAEMVDGLLRIRRAELPG-----EAVAVATGGFSRTV 232

Search completed: June 24, 2003, 22:23:50
Job time: 12.3982 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.30328 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-3
Perfect score: 1257
Sequence: 1 NKRAAFMLLLFLRSVLKVL.....GRRTTSILVATGLAKLIN 250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PTOUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	7.3	463	US-09-134-001C-4160	Sequence 4160, Ap
2	91	7.2	524	US-08-853-659A-35	Sequence 35, Appl
3	84.5	6.7	542	US-08-968-563-16	Sequence 16, Appl
4	84.5	6.7	542	US-08-969-683A-16	Sequence 16, Appl
5	84.5	6.7	542	US-09-297-928-12	Sequence 12, Appl
6	83.5	6.6	764	US-08-177-109A-2	Sequence 2, Appl
7	83.5	6.6	764	US-08-687-706-2	Sequence 2, Appl
8	83	6.6	535	US-09-117-860-18	Sequence 18, Appl
9	82	6.5	398	US-08-641-038A-2	Sequence 2, Appl
10	82	6.5	398	US-09-059-178-2	Sequence 2, Appl
11	82	6.5	398	US-08-642-541-2	Sequence 2, Appl
12	82	6.5	398	US-09-260-889-2	Sequence 2, Appl
13	82	6.5	398	US-09-347-878-12	Sequence 12, Appl
14	82	6.5	398	US-09-479-275-7	Sequence 7, Appl
15	82	6.5	399	US-09-494-921-2	Sequence 2, Appl
16	82	6.5	831	US-09-269-861A-8	Sequence 8, Appl
17	80	6.4	1024	US-09-522-666-2	Sequence 2, Appl
18	79.5	6.3	607	US-09-204-208A-11	Sequence 11, Appl
19	79	6.3	859	US-09-199-637A-281	Sequence 281, App
20	78	6.2	456	US-07-677-413-3	Sequence 3, Appl
21	78	6.2	535	US-09-117-860-24	Sequence 24, Appl
22	77.5	6.2	478	US-09-093-448-1	Sequence 1, Appl
23	77	6.1	330	US-09-134-001C-3582	Sequence 3582, Ap
24	77	6.1	448	US-09-134-001C-4145	Sequence 4145, Ap
25	77	6.1	478	US-09-040-799-3	Sequence 3, Appl
26	77	6.1	478	US-09-093-448-2	Sequence 2, Appl
27	77	6.1	478	US-09-093-448-3	Sequence 3, Appl

28	77	6.1	857	4	US-08-887-534A-47	Sequence 47, Appl
29	76	6.0	300	4	US-09-433-598-2	Sequence 2, Appl
30	76	6.0	312	4	US-09-134-001C-3534	Sequence 3534, Ap
31	76	6.0	559	4	US-09-242-690A-15	Sequence 15, Appl
32	76	6.0	559	4	US-09-298-924-6	Sequence 6, Appl
33	75.5	6.0	313	4	US-09-134-001C-5185	Sequence 5185, Ap
34	75.5	6.0	434	1	US-08-328-256-11	Sequence 11, Appl
35	75.5	6.0	436	2	US-08-307-588-2	Sequence 2, Appl
36	75.5	6.0	496	1	US-08-328-256-12	Sequence 12, Appl
37	75.5	6.0	557	1	US-08-328-256-10	Sequence 10, Appl
38	75.5	6.0	557	1	US-08-471-454-2	Sequence 2, Appl
39	75.5	6.0	557	2	US-08-466-974-2	Sequence 2, Appl
40	75.5	6.0	557	2	US-08-471-453-2	Sequence 2, Appl
41	75.5	6.0	557	2	US-08-307-588-4	Sequence 4, Appl
42	75.5	6.0	756	4	US-09-134-001C-3553	Sequence 3553, Ap
43	75.5	6.0	861	1	US-08-484-105-18	Sequence 18, Appl
44	75.5	6.0	861	1	US-08-484-106-18	Sequence 18, Appl
45	73.5	5.8	256	4	US-09-504-358-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-4160
; Sequence 4160, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4160
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4160

Query Match	7.3%	Score 92;	DB 4;	Length 463;
Best Local Similarity	25.3%	Pred. No. 0.039;		
Matches	60;	Conservative 34;	Mismatches 93;	Indels 50; Gaps 13;
QY	17	KVILVL-----DVGNTNIVLGIYND--TKLTAEWRLSTDVLRSD---	EYGIQVMNL	63
Db	36	KIIIVSAPGKRHNDIKTTDLIRLYEKVLNKLNYESKKOEIIQRYADIVELGINDIL	95	
QY	64	FQQKLDPTLVEGVIISSVVPNMVYSLHEMIRKFKINPLVVGPGIKTGINTKYNPKREV	123	
Db	96	I---TINDTLEBYIKHLSDKPNRLDALLSCGENEFAQ-LIAQYNNQSGIPIRYISPKEA	151	
QY	124	GAD-----RIVNVAHAHEIYK-----RSLIIDF-----GTATTCFCAVRENGDYL	164	
Db	152	GLTVVLDLPOQAQILDS-AYNEIYKLRDYDEKLIIPGFFGVSKONIVTF--PRGSDITG	208	
QY	155	GAICPGIKVSSSALPEKAAPRVELIKPAYAICKNTISS---IQSGIVYRYLRQVKY	219	
Db	209	ALIARGVRAS---LYENETDVSIGIYKANP-----NIINNPILIEIYREMRRELSY	256	

RESULT 2
US-08-853-659A-35
; Sequence 35, Application US/08853659A
; Patent No. 5925522
; GENERAL INFORMATION:
; APPLICANT: Wong, K.K.; Saffer, J.D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods of Detection

```

: RESULT 3
: US-08-968-563-16
: Sequence 16, Application US/08968563
: Patent No. 6013494
: GENERAL INFORMATION:
: APPLICANT: CHARLES E. NAKAMURA
: APPLICANT: ANTHONY A. GATENBY
: APPLICANT: AMY (KUANG-HUA) HSU
: APPLICANT: RICHARD D. LA REAU
: APPLICANT: SHARON L. HAYNIE
: APPLICANT: MARTA DIAZ-TORRES
: APPLICANT: DONALD E. TRIMBUR
: APPLICANT: GREGORY M. WHITED
: APPLICANT: VASANTHA NAGARAJAN
: APPLICANT: MARK S. PAYNE
: APPLICANT: STEPHEN K. PICATAGGIO
:

```

RESULT 4

US-08-969-683A-16
; Sequence 16, Application US/08969683A
; Patent No. 6136576
; GENERAL INFORMATION:
; APPLICANT: GENENCOR INTERNATIONAL, INC.
; TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
; PRODUCTION OF 1,3 PROPANEDIOL
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 4 Cambridge Place
; CITY: Rochester
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 14618
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/969,683A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/20873
; FILING DATE: 13-NOV-1997
; APPLICATION NUMBER: 60/030,601
; FILING DATE: 13-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC 369-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-864-7620
; TELEFAX: 650-845-6504
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: GLPABC
; -08-969-683A-16

Query Match 6.7%; Score 84.5; DB 4; Length 542;
Best Local Similarity 21.0%; Pred. NO. 0.38;
Matches 54; Conservative 31; Mismatches 89; Indels 83; Gaps 11;
QY 24 VGNNTIVLGIYN---DTKLTAEWRLSTDVLSADEYGIQVNMIFQDKLDPTLVGVIIIS 80
Db 266 IGTTSRLID-YNEIDNNRTAE---EVDIL-----LREGEKLAPVMAK----- 304
QY 81 SWPNIMYSLEHMIRKFKINPLVVGPGIKTGINIK-----YDNPKEVGADRIYNAVAH 135
Db 305 -----TRILRAYSGVRPLVASDDPSGRNLSRGVILLDHAERDGLD----- 345
QY 136 EYKRSLIIDFGTATTCFAVRENGDYLGGAI-----PGIKVSSEALF 179
Db 346 -----GFTITGGKLMY---RLMAEWATDAVCKLGNTRCTTADIALPQSQEPAEVTL 397
QY 180 EKAAKLPVELIKPAYAICKNTWISSIQSGIYRYL-----RQVKYLEKLEN--L 228
Db 398 RKVISLPAPLRGSAVYRHGDRTPAWLSEGLRHSILVCEAVTAGEVQYAVENLNVSLL 457
QY 229 PDGRRTRTSLVATGGL 245
Db 458 DLRRRTVGMGTCCQGL 474

RESULT 5
US-09-297-928-12
; Sequence 12, Application US/09297928
; Patent No. 6358716
; GENERAL INFORMATION:
; APPLICANT: BULTHUIS, BEN A.
; GATENBY, ANTHONY A.
; HAYNIE, SHARON L.
; HSU, AMY K.
; LAREAU, RICHARD D.
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF
; GLYCEROL BY RECOMBINANT
; ORGANISMS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 4 CAMBRIDGE PLACE
; CITY: ROCHESTER
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 14618
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/297,928
; FILING DATE: 11-May-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/03602
; FILING DATE: NOVEMBER 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9981-PI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; TELEX: 6717325
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-297-928-12

Query Match 6.7%; Score 84.5; DB 4; Length 542;
Best Local Similarity 21.0%; Pred. NO. 0.38;
Matches 54; Conservative 31; Mismatches 89; Indels 83; Gaps 11;
QY 24 VGNNTIVLGIYN---DTKLTAEWRLSTDVLSADEYGIQVNMIFQDKLDPTLVGVIIIS 80
Db 266 IGTTSRLID-YNEIDNNRTAE---EVDIL-----LREGEKLAPVMAK----- 304
QY 81 SWPNIMYSLEHMIRKFKINPLVVGPGIKTGINIK-----YDNPKEVGADRIYNAVAH 135
Db 305 -----TRILRAYSGVRPLVASDDPSGRNLSRGVILLDHAERDGLD----- 345
QY 136 EYKRSLIIDFGTATTCFAVRENGDYLGGAI-----PGIKVSSEALF 179
Db 346 -----GFTITGGKLMY---RLMAEWATDAVCKLGNTRCTTADIALPQSQEPAEVTL 397
QY 180 EKAAKLPVELIKPAYAICKNTWISSIQSGIYRYL-----RQVKYLEKLEN--L 228
Db 398 RKVISLPAPLRGSAVYRHGDRTPAWLSEGLRHSILVCEAVTAGEVQYAVENLNVSLL 457

Query Match 6.5%; Score 82; DB 2; Length 398;

db 235 RIQGLKNTGAVTSPHDAALIMRGTKTINLRMRHCANAQVIAEELARQPOVELTHYPGI. 29A

Query Match 6.5%; Score 82; DB 2; Length 398;

QY 193 ---PAYAICKNTISSIQSGIVVRYLRQVKYLFEKLENLPDGRTRTSLVL 240
| | : : | : : || : ||| : |
Db 295 ASFPQYTILARQOMSOPGGMIAF-----ELKGIGAGRRFMNALQL 334

```

RESULT 11
US-08-642-541-2
; Sequence 2, Application US/08642541
; Patent No. 5891704
; GENERAL INFORMATION:
; APPLICANT: YUYING, TAN
; TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING
; TITLE OF INVENTION: METHIONINASE, PURIFIED RECOMBINANT METHIONINASE PRODUCED
; TITLE OF INVENTION: METHIONINASE, AND US95 THEREOF
; TITLE OF INVENTION: THEREFROM AND US95 THEREOF

```

Query Match	6.58	Score 82	DB 2	Length 398
Best Local Similarity	22.88	Pred. No. 0.46		
Matches 39	Conservative	25	Mismatches 53	Indels 54
				Gaps 8

Qy 157 RENG--DYLGAICP-----GIK-----VSSEALFEKAAKUPRVELIK---- 192
 | | | | | : | | | | | : : | | : | : | | |
 Db 235 RLQGLKDMTGAVLSPHDAALLMRGIKTNLNMRDHCANAOVLAEFLARQOVELIHY PGL 294

```

QY      193  ---PAYAICKNTISSIQSGIVYRYLRQVKYLFKELKENLPDGRRTTSLVL 240
      | | : : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db      295  ASFPQYTLARQMSQPGGMIAF-----ELKGGIGAGRRFMNALQL 334

```

RESULT 12
US-09-260-889-2
; Sequence 2, Application US/09260889
; Patent No. 6231854
; GENERAL INFORMATION:
; APPLICANT: TAN, Yuying
; TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING
; TITLE OF INVENTION: METHIONINASE

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW, suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA

Query Match 6.5%; Score 82; DB 4; Length 398;
Best Local Similarity 22.8%; Pred. NO. 0.46;
Matches 39; Conservative 25; Mismatches 53; Indels 54; Gaps 8;

QY 157 RENG--DYLGAICP-----GIK-----VSSSEALFEAAKLPRVELIK---- 192

Db 235 RLOGLKDMTGAVLSPHDAALLMRGIKTLNLRMDRHCANAOVLAEELAROBVELIHYPPGL 294

```

193 ---PAYAICKNTISSIQSGIVVRYLRQVKYLFKELKENLPDGRTRTSLVL 240 QY
      | | : : : | | : : | | : | | : | |
295 ASEPOYTLAROOMSOPGGMIAF-----ELKSGIGAGREFMNALOL 334 Db

```

```

RESULT 13
US-09-347-878-12
; Sequence 12, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT OF
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347878
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Pseudomonas putida

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: June 24, 2003, 21:33:11 ; Search time 4.34819 seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453a-3
Perfect score: 1257
Sequence: 1 NKRAAFMLLLFLRSVLKVL.....GRRTRSLVLATGGLAKLIN 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	623	49.6	233	1 YACB_BACSU	P37564 bacillus su
2	141	11.2	267	1 BAF_BORPE	Q4538 bordetella
3	91	7.2	309	1 VGI3_BPT4	P1110 bacteriophage
4	89	7.1	357	1 BUK_THETN	Q87832 thermoanaer
5	88.5	7.0	874	1 CLPA_BRANA	P46523 brassica na
6	88	7.0	586	1 VOLD_BPP2	P13520 bacteriophage
7	86	6.8	280	1 ISPE_CLOAB	Q97f51 clostridium
8	86	6.8	872	1 CLPB_SYNY3	P74361 synchocyst
9	86	6.8	933	1 CC54_YEAST	P30665 saccharomyc
10	85.5	6.8	542	1 GLPA_ECOLI	P13032 escherichia
11	85	6.8	387	1 EX7L_CAMJE	Q9pih4 campylobact
12	85	6.8	926	1 CLAA_LYCPS	P31541 lycopersico
13	84.5	6.7	613	1 YGAR_YEAST	P50080 saccharomyc
14	84	6.7	303	1 V212_FOWPV	Q9j523 fowlpox vir
15	84	6.7	922	1 CLPA_PEA	P35100 pisum sativ
16	84	6.7	3829	1 SACS_HUMAN	Q9nuzj4 homo sapien
17	83.5	6.6	764	1 CFAB_HUMAN	P00751 homo sapien
18	83.5	6.6	856	1 CLPB_HELP1	Q9znh1 helicobacte
19	83	6.6	453	1 EX7L_RICPR	Q9zcp8 rickettsia
20	83	6.6	631	1 YC72_MYCTU	Q11047 mycobacteri
21	82.5	6.6	923	1 CLAB_LYCPS	P31542 lycopersico
22	82.5	6.6	1234	1 RPOB_CLOPE	Q93r88 clostridium
23	82	6.5	209	1 RMJ_ECOLI	P28692 escherichia
24	82	6.5	398	1 MSGL_PSEPU	P13254 pseudomonas
25	82	6.5	532	1 G8PL_BORBU	O51672 borrelia bu
26	82	6.5	856	1 CLPB_HELPY	P17404 helicobacte
27	81.5	6.5	251	1 HXS6_SULSO	O33774 synfolobus
28	81.5	6.5	697	1 PFA_SYNY3	P73662 synchocyst
29	81.5	6.5	761	1 CFAB_MOUSE	P04186 mus musculu
30	81.5	6.5	1081	1 CFAB_SYNY3	O55756 synchocyst
31	81.5	6.5	1122	1 TERT_MOUSE	O70372 mus musculu
32	81	6.4	384	1 CARA_PORPU	P51201 porphyra pu
33	81	6.4	533	1 KPYK_DROME	O62619 drosophila

RESULT 1

YACB_BACSU

ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borcherdt S.,
Bouriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

ALIGNMENTS

34	81	6.4	619	1 HSCA_HAEIN	P44569 haemophilus
35	81	6.4	1159	1 DP3A_PASMU	Q9cpk3 pasteurella
36	80.5	6.4	883	1 CLPB_SYNP7	P53533 synchococc
37	80.5	6.4	1073	1 CARB_METMA	P58944 methanosarc
38	80.5	6.4	1562	1 YMB1_YEAST	Q04781 saccharomyc
39	80	6.4	335	1 RLAA_SULAC	P35023 sulfolobus
40	80	6.4	708	1 HELS_SULSO	Q97vy9 sulfolobus
41	79.5	6.3	274	1 NUCA_ANASP	P38446 anabaena sp
42	79.5	6.3	765	1 BGLX_ECOLI	P33363 escherichia
43	79.5	6.3	1132	1 TERT_OXYTR	O76332 oxytricha t
44	79	6.3	219	1 Y399_BORBU	O51360 borrelia bu
45	79	6.3	336	1 ILVC_THEMA	Q9wz20 thermotoga

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D26185; BAA05305.1; -
CC EMBL; Z99104; CAB11846.1; -
CC Subtilist; Bg10133; YacB.
CC InterPro: IPR004619; Baf.
CC Pfam; PF03309; Bvg_acc_factor; 1.
CC TIGRFAMS; TIGR00671; baf; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;
CC -----
Query Match 49.6%; Score 623; DB 1; Length 233;
Best Local Similarity 55.8%; Pred. No. 9.9e-45;
Matches 116; Conservative 44; Mismatches 48; Indels 0; Gaps 0;
CC -----
Db 18 VILVDVGNNTNVLGIYNDTKLTAEWRLSTVDLRSADVEYGIQVNNLFQODKLDPTLVEGV 77
1-MLLVIDVGNNTNVLGIYHDGKLEYHWRITSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
CC -----
QY 78 IISVVYPNIMYSLEHMIRKYNPLVGVGPKTGKINIKYDNPKEVGADRIYNAVAHEI 137
61 IISVVYPNIMFALERMCTYFHELPQIVGPKMTGLNINIKYDNPKEVGADRIYNAVAIHL 120
CC -----
QY 138 YKRLIIIDFGTATTCVAVRENGDYLGGAIICPGIKVSSSEALFEKAALPRVELIKPAYAI 197
1-MLLVIDVGNNTNVLGIYHDGKLEYHWRITSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db 121 YGNPLVDFGTATTTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETTRDNI 180
CC -----
QY 198 CKNTISSIOSGIVRYLROVKYLFELK 225
181 GRNVSAMSGILFGIVGQGVIVKRMK 208
CC -----
RESULT 2
BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
PC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshaizer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli."
RL J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN-BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a bira homolog in Bordetella pertussis."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN-BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus
```

```
RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis."
RL Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BvgA-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12020; AAA75361.1; -
CC EMBL; AF016461; AAC68834.1; -
CC EMBL; X90711; CAA62242.1; -
CC InterPro: IPR004619; Baf.
CC Pfam; PF03309; Bvg_acc_factor; 1.
CC Transcription regulation; Activator.
CC SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;
CC -----
Query Match 11.2%; Score 141; DB 1; Length 267;
Best Local Similarity 22.5%; Pred. No. 0.00011;
Matches 58; Conservative 43; Mismatches 103; Indels 54; Gaps 8;
CC -----
QY 18 VILVDVGNNTNVLGIYNDTKLTAEWRLSTVDLRSADVEYGIQ 59
1-MLLVIDVGNNTNVLGIYHDGKLEYHWRITSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MIILIDSGNSRLKGVGWFDPDPAQAREPAPVAFNDLDDALGRW-LATLPRRQALGVN 59
CC -----
QY 60 VMNLFOODKLDPTLVEGVIISSVVPNIMYSLEHMIRKYNPLVGVGPKTGKINIKYD 119
60 VAGLARGEIAATLRAG-----GCDIRWLRAQFLAM--GLRNG-----YRN 98
CC -----
QY 120 PREVGADR---IVNAVAHEIYKRSLLIIDFGTATTCVAVRENGDYLGGAIICPGIKVSS 176
1-MLLVIDVGNNTNVLGIYHDGKLEYHWRITSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
Db 99 PDQLGADRWACWGVGLARQPSVHPPLLVASFATGTTLDITGPDNVFPGGLILFGPAMRG 158
CC -----
QY 177 ALFEKAALPRVELIKPAYAI-----CKNTISSIOSGIVRYLROVKYLFELK 226
181 GRNVSAMSGILFGIVGQGVIVKRMK 208
Db 159 ALAYGTAHLPLADGLIVADTPIDTHQAIASIAAQAQAIYQWLQAGRYQQAPEIYVAG 218
CC -----
QY 227 -NLPGGRTRTSLVLTG 243
219 GCMPEVQRQAERLLAVTG 236
Db -----
RESULT 3
VG13_BPT4 STANDARD; PRT; 309 AA.
AC P11110;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neck protein Gpl3.
GN 13.
OS Bacteriophage T4.
PC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B;
RX MEDLINE=89263795; PubMed=2657662;
RA Selivanov N.A., Prilipov A.G., Mesyanzhinov V.V.;
RT "Nucleotide sequences of bacteriophage T4 genes 13, 14 and 15."
RL Nucleic Acids Res. 17:3583-3583(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis."
RN [3]
```

```

CC CC entties requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to licensee@isb-sib.ch).
-----
CC EMBL; AF013165; AAM25356.1; -.
DR DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
DR DR PROSITE; PS01076; ACETATE_KINASE_2; 1.
KW KW transferase; Kinase; Complete proteome.
SQ SEQUENCE 357 AA; 39217 MW; 05CDDE09A2E7EF3F CRC64;

Query Match 7.1%; Score 89; DB 1; Length 357;
Best Local Similarity 20.8%; Pred. No. 3.2;
Matches 57; Conservative 48; Mismatches 75; Indels 94; Gaps 13;

QY QY 16 LKVLVDVGNNTNIVLGVIYNDTK--ITAENRLSTDVL----RSADEYGIQ---VMNLFOQ 66
   | : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db Db 3 LALLILNPNGSTKVAVFRDKEPVPTETLRHSTEELSKYKRIDQFEFTQTALDLMLKE 62

QY QY 67 DKLDPLTVEGVI-----ISSVVPNIMYSL--EH 92
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db Db 63 KGISLSQDAIYVGRLGLKPESGTIVYNKMLEDKKAERGHASNLGAIIATYLAKEH 122

QY QY 93 MIRKYEKINPLVVGP---GIKTGINKYDNPKVEGADRIVNAVAHEIKYR----- 140
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db Db 123 NPAYI-VPVVVDELEDVARITGL-----PEIEKQSIFHALNOKATARRLASDLGKR 174

QY QY 141 ----SLTIIDFGPATTTCFAVRENGDYLGGAICPGIKVSSSEALF--EKAAKLPRLVELIKPA 194
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db Db 175 YDEVNLIIAHLGGGISVGAHRK----GRVIDYNDALNGSGPFSPERAGGLPVLIDLVKLC 229

QY QY 195 YAICKNTWISSIQSVIVRYLRQVKYLFEKLKENL 228
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db Db 230 Y-----SG-----KYTFEEMKKKL 243

RESULT 5
CLPA_BRANA STANDARD; PRT; 874 AA.
AC P46523;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DE ATP-dependent clp protease ATP-binding subunit clpA homolog,
GN chloroplast precursor (Fragment);
DN CLPA.
OC Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosoids II; Brassicales; Brassicaceae; Brassica.
OC NCBI_TaxId=3708;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Topas; TISSUE=Seed;
RX MEDLINE=94218400; PubMed=8165250;
RA Ko K., Duong C., Ko Z.W.;
RL "Nucleotide sequence of a Brassica napus Clp homolog.";
RT Plant Physiol. 104:1087-1089(1994).
CC -!- FUNCTION: MAY INTERACT WITH A CLPP-LIKE PROTEASE INVOLVED IN
CC DEGRADATION OF DENATURED PROTEINS IN THE CHLOROPLAST.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC CC -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC CC -!- SIMILARITY: CONTAINS 1 UVR DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to licensee@isb-sib.ch).
-----
CC EMBL; X75328; CAAS3077.1; -.
CC InterPro; IPR003593; AAA_ATPase.
DR DR InfoPro; IPR003599; AAA_ATPase_Centr.

```


CC step.
CC -!- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE007787; AAK80844.1; -
DR InterPro: IPR001745; GHMPkinse_ATP.
DR InterPro: IPR004424; ISPE.
DR Pfam: PF00288; GHMP_kinases; 1.
DR TIGRFAMs: TIGR00154; ispe; 1.
KW Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
FT NP_BIND 91 101 ATP (POTENTIAL).
SQ SEQUENCE 280 AA; 31304 MW; 42098081A63204F0 CRC64;

Query Match 6.8%; Score 86; DB 1; Length 280;
Best Local Similarity 24.0%; Pred. No. 4.2;
Matches 59; Conservative 31; Mismatches 78; Indels 78; Gaps 16;

QY 36 DTKLTAEWRLSDVL-PSADEXGI-----QVNNLFQODKLDPTLV-EGVISS---VVP- 84
DB 2 NVKAYAKVNSLDVIGKREDGYHLKMTQSNLY--DVLDIRIIDEKIKITSNRRNPT 59

QY 85 ---NTWYSLHMRKFKINPLVGPVIGTGINIKYDNPKEVG-----ADRVNVAVAHE 136
DB 60 NDKNIAYRAAKLFMDTYRID-----KGSIHINKRI--PVAAGLAGGSADGAALVKAMRD 112

QY 137 IYKRSLSL-----LIIDFGTATTCFAVRENGDYLGAICPGIKVSVSEALFEKAALPR-- 187
DB 113 IFPKDVSDEELINLGVKIGADIPFCIVG-----GTAFCEGIG-----EKITKLRSWN 159

QY 188 ---VELIKPAVAICKNTSISSQGIYRYLRQVKYLFKLEKLENLDPGRRTRTSLVATGG 244
DB 160 GRIIVLVKPFQGVSTKMYT-----EYDKCLDVKH-----PDSE-----G 194

QY 245 LAKLIN 250
DB 195 LVKAVN 200

RESULT 8
CLPB_SYNY3 STANDARD; PRT; 872 AA.
AC P74361;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ClpB protein.
GN CLPB OR SLR1641.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Suglura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res 3:109-136(1996).
CC -!- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D90914; BAA18456.1; -
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003959; AAA_ATPase_centrl.
DR InterPro: IPR001270; Chaprnl_cfpA/B.
DR InterPro: IPR004176; Clp_N.
DR Pfam: PF00004; AAA; 2.
DR Pfam: PF02861; Clp_N; 2.
DR PRINTS: PR00300; CLPPROTEASEA.
DR PRODOM: PD000739; GSPIL_E; 1.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00870; CLPAB_1; 1.
DR PROSITE: PS00871; CLPAB_2; 1.
KW Chapterone; ATP-binding; Repeat; Complete proteome.
FT DOMAIN 163 411 I.
FT NP_BIND 537 728 II.
FT NP_BIND 208 215 ATP (POTENTIAL).
FT NP_BIND 611 618 ATP (POTENTIAL).
SQ SEQUENCE 872 AA; 98122 MW; 7EAA8486C7D8D4D5 CRC64;

Query Match 6.8%; Score 86; DB 1; Length 872;
Best Local Similarity 25.2%; Pred. No. 17;
Matches 34; Conservative 20; Mismatches 45; Indels 36; Gaps 6;

QY 52 SADEXGIOMNLFQODKLDPTLVGVISSVVPNIMYSLHMRKFKINPLVVG-PGI- 109
DB 162 SLEKVGRLDELARGKLDVIGRDEEVRRTI-----QILSRTRKNPLVIGEPGVG 213

QY 110 KTGINKYDNPKEVGADRVNVAVAHEIYKRSLLIIDFGTATTCFAVRENGDYLGAICP 169
DB 214 KTAI-----AEGLAQRIINHDPESLDRKLISLDM-----GALTA 249

QY 170 GIKVSEALFEKAAL 184
DB 250 GAKYRGE--FEERLK 262

RESULT 9
CC54_YEAST STANDARD; PRT; 933 AA.
AC P30665;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell division control protein 54.
GN CDC54 OR HCD21 OR YPR019W OR YP9531.13.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Dalton S.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AE972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,

RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.,
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
RN [3]
RP SEQUENCE OF 571-646 FROM N.A.
RX MEDLINE=93087163; PubMed=1454522;
RA Coxon A., Maundrell K., Kearsey S.E.;
RT "Fission yeast cdc21+ belongs to a family of proteins involved in an
RT early step of chromosome replication.";
RL Nucleic Acids Res. 20:5571-5577(1992).
CC -|- FUNCTION: REQUIRED FOR S PHASE EXECUTION.
CC -|- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -|- SIMILARITY: BELONGS TO THE MCM FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U14731; AAA86310.1; -;
DR EMBL; 249919; CAA90164.1; -;
DR EMBL; 271255; CAA95015.1; -;
DR EMBL; 215032; CAA78750.1; -;
DR PIR; S25527; S25527;
DR PIR; S26641; S26641;
DR SGD; S0006223; CDC54.
DR InterPro; IP001208; MCM.
DR Pfam; PF00493; MCM; 1.
DR ProDom; PD001041; MCM; 1.
DR SMART; SM00350; MCM; 1.
DR PROSITE; PS00847; MCM; 1; 1.
DR PROSITE; PS50051; MCM; 2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW DNA replication; Cell cycle; ATP-binding.
FT DOMAIN 518 725 MCM.
FT NP_BIND 568 575 ATP (POTENTIAL).
SQ SEQUENCE 933 AA; 105003 MW; 382D542AE38975E3 CRC64; .

Query Match 6.8%; Score 86; DB 1; Length 933;
Best Local Similarity 17.6%; Pred. No. 18;
Matches 37; Conservative 59; Mismatches 80; Indels 34; Gaps 9;

QY 23 DVGNTNVLGIYNDKLTAEWRLSTDLVLSADEYGLGVNMLFQDQKLDPTLVEGVLIISV 82
DB 235 ELGTSNLN-----DARNLLAYQTELYHQLLNYPQEVISI-----MDQTIKDCMVSLLV 285
QY 83 VPIVNTSLRHMIRKFKINPLVVGPGIKTGINKVDNPKVEGADRVNVAHAHEIYKRL 142
DB 286 DNNLDYDLDEIETKFKVPYVNG-----SCKGMRELNPNDI--DKLNLKGL--VLRSTP 337
QY 143 IIDEFGTATTCVAVRENGYDLGATCGPKGVSEALFEKAALPRVELIKPACIKNTI 202
DB 338 VIPDMKVAFFKNCVCDH-----TMAVEI-DRGVIQPARCERIDCNEP-----NSM 382
QY 203 SSISGIVVRYLRQVLYPEKLEKLENLPDGR 232
DB 383 SLIHNRCSPADKQVIR--LQETPDFVDPDQ 410

RESULT 10
GLPA_ECOLI
ID GLPA_ECOLI STANDARD; PRT; 542 AA.

AC PL3032; P78238;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Anaerobic glycerol-3-phosphate dehydrogenase subunit A (EC 1.1.99.5)
DE (G-3-P dehydrogenase).
GN GLPA OR B2241 OR Z3499 OR ECS3126.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RX STRAIN=K12;
RX MEDLINE=88227815; PubMed=3286606;
RA Cole S.T., Eiglmeyer K., Ahmed S., Honore N., Elmes L., Anderson W.F.,
RA Weiner J.H.;
RT "Nucleotide sequence and gene-polypeptide relationships of the glpABC
RT operon encoding the anaerobic sn-glycerol-3-phosphate dehydrogenase
RT of Escherichia coli K-12.";
RL J. Bacteriol. 170:2448-2456(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -|- FUNCTION: CONVERSION OF GLYCEROL 3-PHOSPHATE TO DIHYDROXYACETONE.
CC -|- USES FUMARATE OR NITRATE AS ELECTRON ACCEPTOR.
CC -|- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + acceptor = glycerone
CC phosphate + reduced acceptor.

CC -1- COFACTOR: FAD AND FMN. FAD IS SAID TO BE ASSOCIATED WITH GLPA.
 CC -1- PATHWAY: FIRST STEP IN THE TRANSFER OF REDUCING EQUIVALENTS
 CC FROM G-3-P TO A SHORT ELECTRON TRANSFER CHAIN TERMINATING WITH
 CC FUMARATE AS THE ULTIMATE ELECTRON ACCEPTOR.
 CC -1- SUBUNIT: COMPOSED OF A CATALYTIC GLPA/B DIMER AND OF MEMBRANE
 CC BOUND GLPC.
 CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE CYTOPLASMIC MEMBRANE
 CC OFTEN OCCURRING IN VESICLES ASSOCIATED WITH FUMARATE REDUCTASE.
 CC -1- SIMILARITY: BELONGS TO THE FAD-DEPENDENT GLYCEROL-3-PHOSPHATE
 CC DEHYDROGENASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M20938; AAA83864.1; -
 CC EMBL; AE000314; AAC75301.1; -
 CC EMBL; D90855; BAA16060.1; -
 CC EMBL; AE005456; AAC57372.1; -
 CC EMBL; AP002561; BAB36549.1; -
 CC PIR; A32006; DEECNA.
 CC EcoGene; EG10391; glpa.
 CC InterPro; IPR000927; D_aa_oxidase.
 CC InterPro; IPR000447; FAD_Gly3p_dh.
 CC InterPro; IPR000205; NAD_binding.
 CC Pfam; PF01266; DAO; 1.
 CC PRINTS; PR01001; FADG3PDH.
 CC PROSITE; PS00977; FAD_G3PDH_1; 1.
 CC PROSITE; PS00978; FAD_G3PDH_2; 1.
 CC Oxidoreductase; Flavoprotein; FAD; Complete proteome.
 KW NP_BIND 10 38 FAD (POTENTIAL).
 FT CONFLICT 329 329 V -> L (IN REF. 1).
 FT SEQUENCE 542 AA; 58958 MW; E5C803F89E912E0E CRC64;
 SQ
 Query Match 6.8%; Score 85.5; DB 1; Length 542;
 Best Local Similarity 21.08; Pred. No. 10;
 Matches 54; Conservative 31; Mismatches 89; Indels 83; Gaps 11;
 QY 24 VGNINVLGIYN---DTKLAEMRLSTDLVLSADEYGIQVMNLFQDKLDPTEIVEGVIIS 80
 DB 266 IGTTSLRID-YNEIDNNKVTAE---EVDIL-----LREGEKLPVMAK----- 304
 QY 81 SVVPMNIMSYLHMIRKFKYKINPLVVGPGIKTGNIK-----YDNPKEVGADRIVNAVAH 135
 DB 305 -----TRILRAYSGVRPLVASDDPSGRNVSRGIVLLDHAERDGLD----- 345
 QY 136 EYKRSLLIIDFGTATTCFAVRENGDYLGGAI-----PGIKVSSSEALF 179
 DB 346 -----GFTITGGKLMY---RLMAEWATDAVCKLGNTRPCTTADLALPGSQPAEVL 397
 QY 180 EKAALPRVELIKPAYAICKNTISSIGIVRYL-----ROVKYLFELKEN--L 228
 DB 398 KVLISPLAPLGSAAVYRHGDRTPAWLSEGLHRSLVCECEAVTAGVQYAVENLVNSLL 457
 QY 229 PDGRRTRTSLVATGGL 245
 DB 458 DLRRTVRVGMGTQOQEL 474
 RESULT 11
 EX7L_CAMJE STANDARD; PRT; 387 AA.
 AC Q9PIH4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
 DE (Exonuclease VII large subunit).
 GN XSEA OR C30325.

OS Campylobacter jejuni.
 OC Bacteria: Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences";
 RL Nature 403:665-668(2000).
 CC -1- FUNCTION: BIDIRECTIONALLY DEGRADES SINGLE-STRANDED DNA INTO LARGE
 CC ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
 CC INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
 CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
 CC -1- SUBUNIT: HETEROOLIGOMER COMPOSED OF LARGE AND SMALL SUBUNITS (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE XSEA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; ALI39074; CAB72792.1; -
 CC InterPro; IPR003753; Exonuc_VII_L.
 CC InterPro; IPR004365; TRNA_anti.
 CC Pfam; PF01336; TRNA_anti; 1.
 CC Pfam; PF02601; Exonuc_VII_L; 1.
 CC TIGRFAMs; TIGR00237; xsea; 1.
 KW Hydrolyase; Nuclease; Exonuclease; Complete proteome.
 SQ SEQUENCE 387 AA; 44088 MW; 10E3FD7D55FE9FB8 CRC64;
 Query Match 6.8%; Score 85; DB 1; Length 387;
 Best Local Similarity 22.08; Pred. No. 7.6;
 Matches 50; Conservative 36; Mismatches 83; Indels 58; Gaps 10;
 QY 4 AAFMLLLRSV-----LKVLVLDVGNINVLGIYNIDTKLTAE-WRLSTDVLSADEYGI 58
 DB 173 APFSLIQAALKADDMDLDVLIARGGGREDLFCFNDENLAREITKAKTPIISAIGHEID 232
 QY 59 QVMNLFQDKLDP-----LVEGVIISSVVPNIMSYLHM-----I 94
 DB 233 YVISDFVADFRAPTPSAIDTLFYSKL--DIEQSLDLMEEKLMOLWNYKIQNYENLLNL 290
 QY 95 RKYFKIN-PLVVGPGIKTGINI-----KYDNPKEVGADRIVNAVAHAETVKR 140
 DB 291 SKPKFNSLPKIIDEKIKQSHNIEKQLNHLLANOMRYN---ELKDLKQNLAYLOHNF-- 345
 QY 141 SLIIDFGTATTCFAVRENGDY--LGGAIKPGIKVSSSEALFEKAAL 185
 DB 346 -----FNKSKKFCIRKNGKIANLEDLKSDDIVLSSQTSQKEAKI 386
 RESULT 12
 CLAA_LYCES STANDARD; PRT; 926 AA.
 ID CLAA_LYCES
 AC P31541;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-dependent c1p protease ATP-binding subunit c1pA homolog CD4A,
 DE chloroplast precursor.


```

Db 407 YNDSAAVHLVACIIITVTVAILCGA--ITDKTRQIKPIIVISSIFGVGVAGILTLIN 464
QY 115 IKYDNPKVEGADRIYNAAHEIYKRSLLI-----IDFGATTCFAVREN-G 160
Db 465 NNANSAHIGL-LILPGVAFGLAQSSMLASQIQDKKSPTRFSVITFTFCRNIG 523
QY 161 DYLGAICPGIKVSEALFEKAA--KLPRVELIKPAYAICKNTISSIQSGIYVR 212
Db 524 QALGGVI-----SNTVESAIAIKLTKANIQLP-----DGTVDNLVIYR 563

RESULT 14
V212_FOWPV
ID V212_FOWPV STANDARD; PRT; 303 AA.
AC Q9J523;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine/threonine protein kinase FPV212 (EC 2.7.1.-).
GN FPV212.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus.";
RL J. Virol. 74:3813-3831(2000).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC POXVIRUSES SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF198100; AAF44556.1; -.
CC HSP; Q06486; ICKI.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC Pfam; PF00069; pkinase; 1
CC ProDom; PD000001; Euk_pkinase; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transferase; Serine/threonine-protein kinase; ATP-binding.
CC FT DOMAIN 25 303 PROTEIN KINASE
CC FT NP_BIND 31 39 ATP (BY SIMILARITY).
CC FT BINDING 54 54 ATP (BY SIMILARITY).
CC FT ACT_SITE 160 160 ATP (BY SIMILARITY).
CC SEQUENCE 303 AA; 34775 MW; 481BFD17C5220716 CRC64;

Query Match 6.7%; Score 84; DB 1; Length 303;
Best Local Similarity 22.4%; Pred. No. 6.8;
Matches 44; Conservative 33; Mismatches 61; Indels 58; Gaps 9;

QY 81 SVPNIMTSLHEMRKYKFNPLVVGVP-GIKTGINKYDNPKP-----VGADRIYNAAVA 134
Db 113 AIIQRLGRDLENILSEKEKFNITVIKLAIKLIDILKFIHSKEFSGDIKAGNILFGKDD 172
QY 135 HEIYKRSLLIIDFGPATTFCVAVRENGDYLGGAICP-----GIKVSSEA 177
Db 173 DKVY-----LVDYGLATKYSS---NGKHKEVTINPKRHNHGTMAFTSIDAHKGVTVSRRG 224
QY 178 LFEK-----AAKLP-----RVELIKPAYA--ICKNTISSIQSGIYVYLRQV 217
Db 225 DLESLGFCMLKWYSGKLPEWKEYEPENVOQMKEAFVNNISKKTIPFPKNAGLIYNIYKV 284

```

```

QY 218 KYL-----PEKLKE 226
Db 285 TKLEYEEAPNYESLKQ 300

RESULT 15
CLPA_PEA
ID CLPA_PEA STANDARD; PRT; 922 AA.
AC P35100;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-dependent clp protease ATP-binding subunit clpA homolog,
DE Chloroplast precursor.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Little Marvel;
RX MEDLINE=93184208; PubMed=8443344;
RA Moore T., Keegstra K.;
RT "Characterization of a cDNA clone encoding a chloroplast-targeted Clp
RT homologue.";
RL Plant Mol. Biol. 21:525-537(1993).
CC -!- FUNCTION: MAY INTERACT WITH A CLPP-LIKE PROTEASE INVOLVED IN
CC DEGRADATION OF DENATURED PROTEINS IN THE CHLOROPLAST.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC -!- SIMILARITY: CONTAINS 1 UVR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L09547; AAA33680.1; -.
CC PIR; S31164; S31164.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_centr.
CC InterPro; IPR001270; Chaprinin_clpA/B.
CC InterPro; IPR004176; Clp_N.
CC InterPro; IPR001943; UVR/C.
CC Pfam; PF00004; AAA; 2.
CC Pfam; PF02151; UVR; 1.
CC Pfam; PF02861; Clp_N; 2.
CC PRINTS; PR00300; CLPPROTEASEA.
CC ProDom; PD000739; GSPIL_E; 1.
CC SMART; SM00382; AAA; 2.
CC PROSITE; PS00870; CLPAB_1; 1.
CC PROSITE; PS00871; CLPAB_2; 1.
CC PROSITE; PS00151; UVR; 1.
CC Chaperone; ATP-binding; Repeat; Chloroplast; Transit peptide.
CC FT TRANSIT 1 7 CHLOROPLAST (POTENTIAL).
CC FT CHAIN 1 922 ATP-DEPENDENT CLP PROTEASE ATP-BINDING
CC SUBUNIT CLPA HOMOLOG.
CC FT DOMAIN 255 502 I.
CC FT DOMAIN 509 544 UVR.
CC FT DOMAIN 569 760 II.
CC FT NP_BIND 300 307 ATP (POTENTIAL).
CC FT NP_BIND 643 650 ATP (POTENTIAL).
CC SEQUENCE 922 AA; 102710 MW; 1CA4C1BA7D453DFE CRC64;

Query Match 6.7%; Score 84; DB 1; Length 922;
Best Local Similarity 23.8%; Pred. No. 26;
Matches 38; Conservative 23; Mismatches 49; Indels 50; Gaps 7;

```

```

QY 24 VGNINVLGYNDTKLTA-----EWRLSTDVLR-----SAD----- 54
Db 193 IGSEHLLGLREGEGVAARVLENLGADPTNIRTQVIRMVGESADSVTATVSGSSNNKT 252
QY 55 -----EYGIQVMNLFQODKLDPTLVEGVISSVVPNIMYSLEHMIRKYFKINPLVWG-PGI 109
Db 253 PTLLEYGTNLTKLAEEGKLDPVVGRQPIERTV-----QILGRRTKNNPCLIGEPGV 304
QY 110 -KTGINIKYDNPKEVGADRVNAAVAHEIYKRSLLIIDFG 148
Db 305 GKTAI-----AEGLAQRIANGDVPETIEGKKVITIDMG 337

```

Search completed: June 24, 2003, 21:51:04
Job time : 6.34819 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 20.3901 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-3
Perfect score: 1257
Sequence: 1 NKRAAFMLLLFLSLVKVL.....GRTRTSLVLATGLAKLIN 250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_viruses.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119.5	89.1	273	16	Q97EB4 clostridium
2	794.5	63.2	259	16	Q8XHL5 clostridium
3	648.5	51.6	254	16	Q9KGH5 bacillus ha
4	639.5	50.9	255	16	Q8R7M2 thermoanaer
5	637.5	50.7	258	2	Q9F985 bacillus st
6	630.5	50.2	259	16	Q92F54 listeria in
7	630.5	50.2	259	16	Q8YAC5 listeria mo
8	498.5	39.7	265	16	Q9X8N6 streptomyc
9	460.5	36.6	261	16	Q9A6Z1 caulobacter
10	389	30.9	272	16	O06282 mycobacteri
11	384.5	30.6	256	16	Q8RPE4 fusobacteri
12	379	30.2	274	16	Q9CD56 mycobacteri
13	337	26.8	274	16	Q9RX54 deinococcus
14	333.5	26.5	246	16	Q9WZ15 thermotoga
15	327.5	26.1	212	2	O32514 desulfovibr
16	282.5	22.5	273	16	O83446 treponema p

17	262	20.8	262	16	O51477 borrelia bu
18	168	13.4	592	16	Q9JXF1 neisseria m
19	168	13.4	592	16	Q9JWF7 neisseria m
20	161	12.8	276	16	Q8YQD7 anabaena sp
21	159.5	12.7	248	16	Q9HWC1 pseudomonas
22	143	11.4	295	16	Q8Y2M4 ralstonia s
23	137.5	10.9	224	16	Q98Q93 mycoplasma
24	123.5	9.8	229	16	O67753 aquifex aeo
25	121.5	9.7	257	16	P74045 synechocyst
26	114	9.1	56	2	P94305 bacillus ps
27	110.5	8.8	223	16	Q92KF6 helicobacte
28	108.5	8.6	242	16	Q9PC14 xylella fas
29	103.5	8.2	460	2	Q93HV8 arthrobacte
30	102.5	8.2	223	16	O25533 helicobacte
31	99	7.9	536	16	O8XQG4 ralstonia s
32	99	7.9	641	16	Q92FA2 helicobacte
33	98	7.8	874	10	O81723 arabidopsi
34	98	7.8	964	10	Q8VYJ7 streptococ
35	97	7.7	322	2	Q8VVB8 listeria mo
36	96.5	7.7	608	16	O8Y742 synechococ
37	94.5	7.5	895	2	O34209 synechococ
38	93.5	7.4	608	16	Q92BM4 listeria in
39	92.5	7.4	1006	16	O67588 aquifex aeo
40	92	7.3	355	17	Q8U1L9 pyrococcus
41	91.5	7.3	453	16	Q928Q9 listeria in
42	91	7.2	408	16	O8R9H1 thermoanaer
43	91	7.2	524	2	O85311 salmonella
44	90.5	7.2	180	16	Q92NQ7 rhizobium m
45	90.5	7.2	898	16	P74459 synechocyst

ALIGNMENTS

RESULT 1

Q97EB4	ID	Q97EB4	PRELIMINARY;	PRT;	273 AA.
AC	Q97EB4;				
DT	01-OCR-2001 (TREMBLrel. 18, Created)				
DT	01-OCR-2001 (TREMBLrel. 18, Last sequence update)				
DE	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Predicted transcriptional regulator, homolog of Bvg accessory factor.				
GN	CAC3200.				
OS	Clostridium acetobutylicum.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;				
OC	Clostridiales; Clostridiaceae; Clostridium.				
OX	NCBI_TaxID=1488;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 824 / DSM 792 / VKM B-1787;				
RX	MEDLINE=21359325; PubMed=11466286;				
RA	Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;				
RA	"Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."				
RL	J. Bacteriol. 183:4823-4838(2001).				
DR	EMBL; AE007815; AAK81136.1; -				
DR	InterPro; IPR004619; Baf.				
DR	InterPro; IPR000515; BPD.transp.				
DR	Pfam; PF03309; Bvg_acc.factor; 1.				
DR	TIGRFAMs; TIGR00671; Baf; 1.				
DR	PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; UNKNOWN_1.				
KW	Complete proteome.				
SQ	SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;				

Query Match 89.1%; Score 1119.5; DB 16; Length 273;
Best Local Similarity 91.8%; Pred. No. 2.9e-88;
Matches 224; Conservative 9; Mismatches 10; Indels 1; Gaps 1;
OY 7 MLLFLSLVKVLVDVGNINVLGIYNDTKLTAEWRLSTDLRSADVGIVMNLFOQ 66

DR EMBL; AE013180; AM25520.1; --
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBCA86D CRC64;

Query Match 50.9%; Score 639.5; DB 16; Length 255;
Best Local Similarity 53.0%; Pred. No. 4.9e-47;
Matches 123; Conservative 48; Mismatches 56; Indels 5; Gaps 1;

QY	18	VILVDVGNNTNVLGIYNDTKLTAEWRLSTDLVSADDEYGIQVMNLFQQDKLPDPTLVGVG	77
Dd	1	MLLAFDVGNTNVMGVKPKLLHSFRISDTKNKYDEYGMVNLQLIGYNGISLTDIDVV	60
QY	78	IISVVVPNMYSLEHMIRKFKINPLVVGPGIKTGINKYDNKPKEVGADRIVNAVAHEI	137
Dd	61	IISVVPPMLMTLOVMSLKYPTRPIVVGPGIKTGINKYDNKPKEVGADRIVNAVAAYEL	120
QY	138	KRSLIIIDFGTATTCFAVRNGDYLGACPCPGIKVSSEALFEKAALKPRVELIKPAYAI	197
Dd	121	YGGPVIVDFGTATTCAISEKGELGGIAPGLMISADALFORTAKLPKDITKRPPTVI	180
QY	198	CNTSISSQSGIVRYLRQVKYLEKLKENLPDGRRTTSLSLATGGLAKLI	249
Dd	181	NRNTVASMQSGLIYGHVGDVYIVTRMK-----GEFAPSAYVATGCFANMI	227

RESULT 5
Q9F985 PRELIMINARY; PRT; 258 AA.

ID	Q9F985	PRELIMINARY; PRT; 258 AA.
AC	Q9F985	PRELIMINARY; PRT; 258 AA.
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DE	Hypothetical protein.	
OS	Bacillus stearothermophilus.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;	
OX	NCBI_TaxID=1422;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-V;	
RA	Vasquez C., Pichuanes S., Saavedra C.; Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.	
RL	EMBL; AF198621; AAG28531.1; --	
DR	InterPro; IPR004619; Baf.	
DR	Pfam; PF03309; Bvg_acc_factor; 1.	
DR	TIGRFAMS; TIGR00671; baf; 1.	
SQ	SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;	

Query Match 50.7%; Score 637.5; DB 2; Length 258;
Best Local Similarity 53.9%; Pred. No. 7.4e-47;
Matches 125; Conservative 43; Mismatches 59; Indels 5; Gaps 1;

QY	18	VILVDVGNNTNVLGIYNDTKLTAEWRLSTDLVSADDEYGIQVMNLFQQDKLPDPTLVGVG	77
Dd	1	MIFLDVGNNTNVLGVYDGDELKHWRHRIETSRGKTDEYGMTKALLNHVGLQFSIDGI	60
QY	78	IISVVVPNMYSLEHMIRKFKINPLVVGPGIKTGINKYDNKPKEVGADRIVNAVAHEI	137
Dd	61	IISVVPPIMALEMRCLKYPHIPIVVGPGIKTGINKYDNKPKEVGADRIVNAVAGIHL	120
QY	138	KRSLIIIDFGTATTCFAVRNGDYLGACPCPGIKVSSEALFEKAALKPRVELIKPAYAI	197
Dd	121	YGSPILIIVDFGTATTCYINEHQYMGAGAPGIMISTEALFARAALKPRIETARPDDII	180
QY	198	CNTSISSQSGIVRYLRQVKYLEKLKENLPDGRRTTSLSLATGGLAKLI	249
Dd	181	GKNTVSAMQAGILYGVQGVNEGIVSRMKASPVPPK-----VIATGGLASLI	227

RESULT 6
Q92F54 PRELIMINARY; PRT; 259 AA.

ID	Q92F54	PRELIMINARY; PRT; 259 AA.
AC	Q92F54	PRELIMINARY; PRT; 259 AA.
DT	01-MAR-2002 (TrEMBLrel. 20, Created)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DE	Hypothetical protein lmo0221.	
OS	Listeria monocytogenes.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;	
OX	NCBI_TaxID=1639;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-EGD-E / SEROVAR 1/2a;	
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Anand A., MEDLINE=21537279; PubMed=11679669;	

[illegible]

```
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordieik G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluerer T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1; -.
DR MEROPS; M41.009; -.
DR L1stIst; LMO00221; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 50.2%; Score 630.5; DB 16; Length 259;
Best Local Similarity 51.3%; Pred. No. 3e-46;
Matches 119; Conservative 54; Mismatches 54; Indels 5; Gaps 1;

QY 18 VILVDVGNINVLGIYNDTKLTAEWRLSTDLVLSADEYGIQVMNLFQDQKLDPTLVEGV 77
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 78 IISVVPNTMYSLEHMIRYFKINPLVPGVGKGTINIKYDNPKEVGADRIVNAVAHEI 137
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 61 IISVVPPIHAMETMCVRYFNIRPLVPGVGKGTINLKNLVDNPREIGSDRIVNAVAASEE 120
QY 138 YKRSLLIIDFGTATTCVAVRENDYLGGAICPGIKVSSSEALFEKAALPRVELIKPAYAI 197
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 121 YGTPVIVDFGATTCYIDESGVYGGALAPGIMISTALYNRAAKLPRVDIAESSQII 180
QY 198 KNTISSIOSGIVYRLQVKYLFKEKLKENLPDGRTRTSVLATGGLAKLI 249
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 181 GKSTSSVMQAGIFGVFGVCGEGIAEMKK- ----QSNAPVVVATGGLARMI 227
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:

RESULT 8
Q9X8N6 PRELIMINARY; PRT; 265 AA.
ID Q9X8N6 AC Q9X8N6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RL Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RL Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
```

```
RL Mol. Microbiol. 21:77-96(1996).
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser H., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL049628; CAB40880.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 39.7%; Score 498.5; DB 16; Length 265;
Best Local Similarity 43.3%; Pred. No. 6.8e-35;
Matches 104; Conservative 50; Mismatches 73; Indels 13; Gaps 4;

QY 18 VILVDVGNINVLGIYNDTKLTAEWRLSTDLVLSADEYGIQVMNLFQDQKLDPTL- --- 73
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 74 ---VEGVISVVVPMYSLHMKRYFKINPLV-VGPGIKGTINIKYDNPKEVGADRIV 129
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 58 GGDIGDIAICATVPVSVLHELRETVRYGDPVAVLVEGKVGTPILTDHPKEVGADRII 117
QY 130 NAVAHEIYKRSLLIIDFGTATTCVAVRENDYLGGAICPGIKVSSSEALFEKAALPRVE 189
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 118 NAVAARELYGGPAIVVDFGTATTFDAVSARGEYIGVIAPIGIEISVEALGVKAQLRKIE 177
QY 190 LKPAVACIKNTISSIOSGIVYRLQVKYLFKEKLKENLPDGRTRTSVLATGGLAKLI 249
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:
QY 178 VARPSRVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARLADDDPDDVT--VIATGGLAPMV 235
Db :|||:||||| :|||: || :|||: || :|||: || :|||: || :|||: || :|||:

RESULT 9
Q9A6Z1 PRELIMINARY; PRT; 261 AA.
ID Q9A6Z1 AC Q9A6Z1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
```

DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS: TIGR00671; Baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 36.68; Score 460.5; DB 16; Length 261;
Best Local Similarity 39.58; Pred. No. 1.2e-31;
Matches 9; Conservative 8; Mismatches 81; Indels 7; Gaps 2;

QY 18 VILVDVGNVTNIVLGIYNDTKLTAEWRLSTDVLKRSADYEYGIQVWNLFOQDKLDPTLVGV 77
DB 2 MLLAIEGNTNTWFAIHGASWVAQWSATSTETADEYVVLWQLLSQGLGFRADAV 61
QY 78 IISVVPNIMYSLEHMTIKYFKINPLVVGPIKGTGINKIKYDNPKEVGADRVNVAHEI 137
DB 62 IISVVPQSTFNLRNLRRRYFNVEPLVIGENAKLIDVRIEKPSEAGADRVNAIGAMV 121
QY 138 YKRSLLIIDGTGTTTCFCAVRENGDYLGCATCGIKVKSSEALFEKAALKPRVELLKPA--Y 195
DB 122 YGPLVVIDSGTATTFDVAAGAFEGITAPGINSQALHEAAKLPRIATORPAGNR 181
QY 196 AICKNTISSIQSGIVYRLQVKYLFKKLENLPDGRTRTSLVLATGGLAKL 248
DB 182 IGVGTDVTSAMQSGVFWGYISLIEGLVARIK-----AERGEPTVIATGGVASL 229

RESULT 10
O06282 PRELIMINARY; PRT; 272 AA.
AC O06282;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 29.3 kDa protein (transcriptional activator, putative, Baf family).
GN RV3600C OR WTCY07H7B.22 OR MT3706.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
RX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagsis K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence";
RT Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z95557; CAB08944.1;
DR EMBL: AE007170; AAK48063.1;
DR TIGR: MT3706;
DR TubercuList; Rv3600c;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS: TIGR00671; Baf; 1.

QW	Hypothetical protein; Complete proteome.	QW	SEQUENCE	272 AA; 29304 MW; 5D70E6EF0F09AC8B CRC64;
Query Match	30.9%; Score 389; DB 16; Length 272;			
Best Local Similarity	34.2%; Pred. No. 1.8e-25;			
Matches	81; Conservative 63; Mismatches 85; Indels 8; Gaps 3;			
QY	18	VILVLDVGNVNIIVLGIY----	DTKLTAEWRLSTDVLRSADEYGIQVNNLFQODKLDPTL 73	
DB	1	MLLAIDVRNTHTVVGLLSGMKEHAKVQVQWRIITSEVTADELALTIDGLIGDS----	ER 57	
QY	74	VEGVITSSVVPNTMYSLEHMIRKYEKINP-LVVGPGIKTKINIKYDNPKEVGADRVNAV 132		
DB	58	LTGTAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDNPNKEVGADRVNCL 117		
QY	133	AAHEIYKRSLLIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIK 192		
DB	118	AAVDFRKAIAVVDGSGSICVDVYSARKEFLGGAIPAGVQVSSDAAAARSAAALRRVELAR 177		
QY	193	PVAICAICTISSIQSGIVRYRLQVKYLFKEKLENLPDGRRTTSLVLTATGGGLAKLI 249		
DB	178	PRSVGANTVECHQAGAVFGAGLVGLGRIEDVSGFSDVDHVAIVATGHTAPLL 234		
RESULT 11				
Q8RFE4	PRELIMINARY; PRT; 256 AA.			
ID	Q8RFE4			
AC	Q8RFE4;			
DT	01-JUN-2002 (TReMBLrel. 21, Created)			
DT	01-JUN-2002 (TReMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)			
DE	Bvg accessory factor.			
GN	FN0761.			
OS	Fusobacterium nucleatum (subsp. nucleatum).			
OC	Bacteria; Fusobacteria; Fusobacterium.			
OX	NCBI_TaxID=76856;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 25586;			
RC	MEDLINE=21886394; PubMed=11889109;			
RA	Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,			
RA	Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,			
RA	Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,			
RA	Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,			
RA	Fonstein M., Kyripides N., Overbeek R.,			
RT	"Genome sequence and analysis of the oral bacterium Fusobacterium			
RT	nucleatum strain ATCC 25586".			
RL	J. Bacteriol. 184:2005-2018(2002).			
DR	EMBL; AEO10586; AAL94957.1; -.			
KW	Complete proteome.			
QW	SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;			
Query Match	30.6%; Score 384.5; DB 16; Length 256;			
Best Local Similarity	37.7%; Pred. No. 4.1e-25;			
Matches	89; Conservative 52; Mismatches 86; Indels 9; Gaps 5;			
QY	18	VILVLDVGNVNIIVLGIY-NDTKLTAEWRLSTDVLRSADEYGIQVNNLFQODKLDPTLVEG 76		
DB	1	MIIGIDGNTHVTGIVDNGNGLTSTRIATNDKMTDEYFSFNNTIKYNEISKVDA 60		
QY	77	VITSSVVPNTMYSLEHMIRKYEKINPLVVGPIKTKINI-KYDNPKEVGADRVNAVAAH 135		
DB	61	ILISSVVPNTMIIITPQFFARKYFKVEATVLDLEKKLPPTFAKNGINTYGFADRIIDTEAM 120		
QY	136	EIY-KRSLIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPA 194		
DB	121	QKYPDKNLVDFDGTATTY-DVLKKGVIYIGGILPGIDMSINALYGNATKLPVKFTTFS 179		
QY	195	YATCKNTISSIQSGIVRYRLQVKYLFKEKLENLPDGRRTTSLVLTATGGGLAKLIN 250		
DB	180	SVLGTDFPMQIAAIFFGYAGQIKHIKKNEEL-----NEEFVLATGGGLKLS 230		

```
RESULT 12
Q9CDS56
ID Q9CDS56 PRELIMINARY; PRT; 274 AA.
AC Q9CDS56;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Hypothetical protein M0232.
GN M0232.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL583917; CAC29740.1; -.
DR Lepronia; M0232; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 30.2%; Score 379; DB 16; Length 274;
Best Local Similarity 34.3%; Pred. No. 1.3e-24;
Matches 83; Conservative 62; Mismatches 81; Indels 16; Gaps 5;

QY 18 VILVDVGNINVLGIYNDT---KLTAEWRLSTDLRSADYGIQVMNLFQDDKLDPTLV 73
Db 1 MLADIVRNTHVTVGLSSKEHAKVQWQRIKTESEVDALEALIIDGLIGDSDS---ER 57
QY 74 VEGVISSVVPNIMYSLHMKRYKINP-LVVGPGIKTGINKYNDPKEVGAADRVNAV 132
Db 58 LAGAAALSTVPSVLHVRIMLDQWPSVPHVLIPEGVRGIPLLVDNPKVGAADRVNCL 117
QY 133 AAHEIKRSLIIDFGTATTCFAVRNGDYLGGAICPGIKVSSEALFEKAALPRVELIK 192
Db 118 AAFKFGQAAIVVDFGSSICVDVWSAKGEFLGAIAPGVQVSSDAASAARLRVELAR 177
QY 193 PAVAICKNTISSQSGIVRYLRQVYLFKLEKLENLPD----GRTRTSLVLTATGLAK 247
Db 178 PRSVGKNIVCEMGAGVVFAGLVGLVRMRQDVEEFSGDLGNRV---AVVATGHTAP 234
QY 248 LI 249
Db 235 LL 236

RESULT 13
Q9RX54
ID Q9RX54 PRELIMINARY; PRT; 262 AA.
AC Q9RX54;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Hypothetical protein DR0461.
GN DR0461.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
```

```
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Kachrova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001905; AAF10040.1; -.
DR TIGR; DR0461; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 26.8%; Score 337; DB 16; Length 262;
Best Local Similarity 33.3%; Pred. No. 5.1e-21;
Matches 80; Conservative 49; Mismatches 89; Indels 22; Gaps 6;

QY 19 ILVLDVGNINVLGIYNDT-KLTAEWRLSTDLRSADYGIQVMNLFQDDKLDPTLV--- 74
Db 6 LLAVDIGNTTTLGLADASGALHTWRTNREMLPDDLALQLHLGLF-----TLGAP 58
QY 75 --EGVISSVVPNIMYSLHMKRYKINP-LVVGPGIKTGINKYNDPKEVGAADRVNAV 132
Db 59 IPRAAVLSSVAPPVGVNGENYALAKRHFMDAFVSAENLPDVTVELTPGSGVADRLCNLF 118
QY 133 AAHEIKRSLI--IIIDFGTATTCFAVRNGDYLGGAICPGIKVSSEALFEKAALPRVE 189
Db 119 GA-EKYLGLDYAVVDFGTSINFDVVGRRFLGILATGAGVSADALFAAAKLPRIT 177
QY 190 LKPAVACKNTISSQSGIVRYLRQVYLFKLEKLENLPDGRTRTSLVLTATGLAKLI 249
Db 178 LQAPETAIGKNTVHALQSLVFGYAEVMDGLLRIRRAELPG-----EAVAVATGFSRTV 232

RESULT 14
Q9WZY5
ID Q9WZY5 PRELIMINARY; PRT; 246 AA.
AC Q9WZY5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
GN TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher C.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -.
DR TIGR; TM0883; -.
```

DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;
Query Match 26.5%; Score 333.5; DB 16; Length 246;
Best Local Similarity 39.1%; Pred. No. 9.4e-21;
Matches 95; Conservative 39; Mismatches 72; Indels 37; Gaps 11;
QY 20 LVLVDGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQODKLDPTL----- 73
DB 3 LLDVGNTHSVFSTEDGKTRRWRLSTGVFQTEDE-----LFSH--LHPLLGDAMRE 53
QY 74 VEGVIISVVPNIMYSLEHMIRKYFKINPLVVGPIKGTGI---NIKVDNPKVEGADRIVN 130
DB 54 IKGIGVASVPTQNTVIERFSQYFHSPIWV--KAKGCVKWNVK--NPSEVGADRIVAN 109
QY 131 AVAAHEIYKRLIIIDFGTATTCFAVRNGDYLGGATCPGKIVSSEALFEKAAKLPVEL 190
DB 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVE- 167
QY 191 IKPA-YAICKNTIISQSGIVRYRLOVKYLFEKLKE---NLPDGRTRTSLVLATGGLA 246
DB 168 VKPADFVVGKDTEENIRLGVVNGSVYALEGIIRIKEVYGDLP-----VVLTTGGQS 218
QY 247 KLI 249
DB 219 KIV 221
RESULT 15
O32514 PRELIMINARY; PRT; 212 AA.
AC O32514;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Similar to Bacillus subtilis.
OS Desulfovibrio vulgaris (strain Miyazaki).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MIYAZAKI;
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,
Kohno K., Akutsu H., Kumagai I., Nakaya T.;
"Sequence analyses of two ferredoxin genes and their flanking regions
from Desulfovibrio vulgaris (Miyazaki F).";
J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).
RL EMBL; AB005550; BAA21476.1; -
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
SQ SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;
Query Match 26.1%; Score 327.5; DB 2; Length 212;
Best Local Similarity 37.7%; Pred. No. 2.5e-20;
Matches 75; Conservative 42; Mismatches 75; Indels 7; Gaps 5;
QY 20 LVLVDGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQODKLDPTLVEGVII 79
DB 6 LLDIGNTNKIGIAVETAVLTSVLPDTPGQTTDSIGRLLEVLRHAGLGPADVGACVA 65
QY 80 SSVVPNIMYSLEHMIRKYFKINPLVVGPG-IKTGINKYDNPKVEGADRIVNAVAHEIY 138
DB 66 SSVVPNPLIRACERYL-YRKLFLAPGDIAIPLDNRYERPAEVGADRLVAAYAAARLY 124
QY 139 --KRLIIDFGTATTCFAVRNGDYLGGATCPGKIVSSEALFEKAAKLPV--ELIKPA 194
DB 125 PGPSRLSVDFGTATTCV-EGGAYLGLGICPGVLSAGALSRTAKLPRLSLEVEEDS 183
QY 195 YAICKNTIISQSGIVRY 213

Db 184 PVIGRSTTSLNHHGFIRGF 202
Search completed: June 24, 2003, 21:59:14
Job time : 21.3901 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.11854 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-3

Perfect score: 1257

Sequence: 1 NKRAAFMLLLFLRSVLKVL.....GRRTRTSLVLTGGLAKLIN 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1119.5	89.1	273	2 E97293	probable transcrip
2	648.5	51.6	254	2 F83660	hypothetical prote
3	630.5	50.2	259	2 AF1102	conserved hypothet
4	630.5	50.2	259	2 AF1464	conserved hypothet
5	623	49.6	233	2 S66100	conserved hypothet
6	498.5	39.7	265	2 T36391	hypothetical prote
7	460.5	36.6	261	2 B87489	transcription acti
8	389	30.9	272	2 A70955	hypothetical prote
9	379	30.2	274	2 H86937	conserved hypothet
10	337	26.8	262	2 E75516	conserved hypothet
11	333.5	26.5	246	2 D73220	conserved hypothet
12	282.5	22.5	273	2 D71326	conserved hypothet
13	262	20.8	262	2 F70165	conserved hypothet
14	168	13.4	592	2 B81009	Bira protein/Bvg a
15	168	13.4	592	2 H82031	probable biotin-[a
16	161	12.8	276	2 A12292	hypothetical prote
17	159.5	12.7	248	2 H83111	hypothetical prote
18	141	11.2	267	2 I40327	baf protein - Bord
19	137.5	10.9	224	2 A99571	conserved hypothet
20	123.5	9.8	229	2 E70465	hypothetical prote
21	121.5	9.7	257	2 S75559	hypothetical prote
22	110.5	8.8	223	2 G71887	hypothetical prote
23	108.5	8.6	242	2 A82637	conserved hypothet
24	102.5	8.2	223	2 F64627	hypothetical prote
25	99	7.9	641	2 F71810	type III DNA modif
26	98	7.8	874	2 G84644	probable ATP-depen
27	96.5	7.7	608	2 AG1259	GTP-binding protei
28	93.5	7.4	608	2 A11621	GTP-binding protei
29	92.5	7.4	1006	2 C70445	ATPase subunit of

30	91.5	7.3	453	2 AD1741	aspartate kinase h
31	91	7.2	309	1 GTBPT4	gene 13 protein -
32	91	7.2	524	2 T14870	hypothetical prote
33	90.5	7.2	898	2 S76431	endopeptidase Clp
34	90.5	7.2	1057	2 T15720	hypothetical prote
35	90	7.2	813	2 D64527	hypothetical prote
36	89.5	7.1	741	2 D75500	ATP-dependent Clp
37	89	7.1	916	2 C75175	hypothetical prote
38	88.5	7.0	635	2 G86744	hypothetical prote
39	88.5	7.0	874	2 S37557	endopeptidase Clp
40	88	7.0	586	2 J00184	old protein - phag
41	88	7.0	1232	2 D64413	cobalamin biosynth
42	87.5	7.0	453	2 AF1371	aspartate kinase h
43	87	6.9	441	2 A71677	tail-specific prot
44	87	6.9	620	2 F70439	oxaloacetate decar
45	86.5	6.9	542	2 H85863	hypothetical prote

ALIGNMENTS

RESULT 1

E97293

probable transcription regulator, homolog of Bvg accessory factor [Imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97293
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97293
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-273 <KUR>

A:Cross-references: GB:AB001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3200
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc.

Query Match	89.1%	Score	1119.5	DB 2	Length	273			
Best Local Similarity	91.8%	Pred. No.	3.9e-88						
Matches	224	Conservative	9	Mismatches	10	Indels	1	Gaps	1
QY	7	MLLFLRSVLKVIILVDVNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVNNLFQQ	66						
Db	1	MLLFLRSVLKVIILVDVNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVNNLFQQ	60						
QY	67	DKLDPTLVEGVLISSVVPNIMYSLEHMIRYFKINPLVVGPGIKTGINKIYDNPKPEVGAD	126						
Db	61	DKLDPTLVEGVLISSVVPNIMYSLEHMIRYFKINPLVVGPGIKTGINKIYDNPKPEVGAD	120						
QY	127	RIVNAVAHEIYKRSLLIIDFGTATTCFCAVRENGDYLGGATCPGKIVKSSSEALFEKAALP	186						
Db	121	RIVNAVAHEIYKRSLLIIDFGTATTCFCAVRENGDYLGGATCPGKIVKSSSEALFEKAALP	180						
QY	187	RVELIKPAYAICAKNTISSIQSGIVYRQVKYLFKEKLENLPDGRTRTSLVLATGCLA	246						
Db	131	RVELIKPAYAICAKNTISSIQSGIVYRQVKYLFKEKLENLPDGRTRTSLVLATGCLA	239						
QY	247	KLIN	250						
Db	240	KLIS	243						

RESULT 2

F83660

hypothetical protein BH086 [Imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83660

R; Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirata,
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphillic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: F83660
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:CROSS-references: GB:AP001507; GB:BA000004; MID:g10172612; PIDN:BAB03805.1; GSPDB:GN00177
A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94_31c

Query Match 51.6%; Score 648.5; DB 2; Length 254;
Best Local Similarity 53.6%; Pred. No. 6.1e+48;
Matches 126; Conservative 41; Mismatches 57; Indels 11; Gaps 2;

Qy 18 VIILVDGNTINVLGIYNDTKLTAEWRLSTDVLSADEYGIQVMNLFOODKLDPTLVGGV 77
 :|||:|||:|||:: |||:: :: |||:: |||: | |||: |||:
 1 MILVIDGNTIVLGYYODETLVHHWRLATSRKTDEYAMTVRSFDHAGLFQDIDGI 60

 78 IISVVVPNIWSLEHMIRKYFKIPLNVVGPGIKTGINKINYDNPKVEYGADRIVNAVAHEI 137
 :|||:|||:|||:: |||:: :: |||:: |||: | |||: |||:
 61 VISSVPPMFMFSLEQMCKKYFHTPMIIIGPIGTGLNIKYPDNPKEVGADRVNAVAAIEL 120

 138 YKRSLIIDFGTATTCAVRENGDYLGGAICPGIKVSSEALFEKAALKRVELIKPAYAI 197
 :|||:|||:|||:: |||:: |||: |||: |||: |||: |||: |||:
 121 YGPALVDFGTATTCTLINEKKQYAGGVIAPIGMISTEALYHRASKLPRIETAKPRQVV 180

 198 CKNTISSIQSGVIYVRLOVKYLFEKLK---ENLPDGRRTRTSVLVLTATGSLAKLI 249
 |||:|||:|||:: |||:: |||: |||: |||: |||: |||: |||:
 181 GTWTIDSMSQSIFGYGVSVQDVGVKRMQAASEPK-----VIATGSLAKLI 227

RESULT 3

AFL102 conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1102
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgeat, O.; Entian, K.D.; Fsihl, H.; Jones, L.N.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makarewicz, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund, O.K.; Schleuter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund, O.
A>Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1102
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <GB>
A:CROSS-references: GB:N_C003210; PIDN:CAD00748.1; PID:gi6409586; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
C:Gene: lmo0221
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94_31c

Query Match 50.2%; Score 630.5; DB 2; Length 259;
Best Local Similarity 51.3%; Pred. No. 2.2e+46;
Matches 119; Conservative 54; Mismatches 54; Indels 5; Gaps 1;

Qy 18 VIILVDGNTINVLGIYNDTKLTAEWRLSTDVLSADEYGIQVMNLFOODKLDPTLVGGV 77
 :|||:|||:|||:: |||:: |||: |||: |||: |||: |||: |||:
 1 MILVIDGNTITCVGYEKQLLKHWRTTDTRHSDELGMTVINFFSANLTPSDIOGI 60

 78 IISVVVPNIWSLEHMIRKYFKIPLNVVGPGIKTGINKINYDNPKVEYGADRIVNAVAHEI 137
 :|||:|||:|||:: |||:: |||: |||: |||: |||: |||: |||:
 61 VISSVPPIMHAMETMCVRIFYFNLRIPLVGPGIKTGUNKLVDPNPREIGSRIYNVAASSEE 120

 138 YKRSLIIDFGTATTCAVRENGDYLGGAICPGIKVSSEALFEKAALKRVELIKPAYAI 197
 :|||:|||:|||:: |||:: |||: |||: |||: |||: |||: |||:

[illegible]

RESULT 7

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.9%; Score 389; DB 2; Length 272;
Best Local Similarity 34.2%; Pred. No. 9.9e-26;
Matches 81; Conservative 63; Mismatches 85; Indels 8; Gaps 3;

QY 18 VILVDVGNNTNIVLGIYNDT---DTKTAEWRLSTDLVRSADYGIQVMNLFQODKLDPTL 73
Db 1 MLLAIDVRNTHVTWGLLSGSKHAKVVOQWRIRTESEVTADELALIDGLIGDS---ER 57
QY 74 VEGVISSVVPNIMYSLEHMIRKYPKINP-LVVGPGIKTGINKYDNPKEVGADRVNAV 132
Db 58 LTTAALSTVPSVLEHVRIMLDQYWPSPVHLIEPGVGTGIPLLVDNPKVEVGADRVNCL 117
QY 133 AAHEIYKRSLLIIDFTATTTCFAVRENGDYLGAICPGIKVSSEALFEKAALPRVELIK 192
Db 118 AAYDRFKAIVVDFGSSICVDVVSAGFELGAIAPGVQVSDAAARSALRRVELAR 177
QY 193 PAYACKNTISSIQSGIVRYRQVKYLFKELKENLPDGRRTTSVLVATGGLAKLI 249
Db 178 PRSVGKNTVECMQAGAVFGFAGLVGLVGRIRVEDVSGFSDVDHDAIVATGHTAPLL 234

RESULT 9

H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H86937
R;Cole, S.T.; Eigemeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho-
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; So-
A;Title: Massive gene decay in the leprosy bacillus
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H86937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0232
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.2%; Score 379; DB 2; Length 274;
Best Local Similarity 34.3%; Pred. No. 7.1e-25;
Matches 83; Conservative 62; Mismatches 81; Indels 16; Gaps 5;

QY 18 VILVDVGNNTNIVLGIYNDT---KLTAEWRLSTDLVRSADYGIQVMNLFQODKLDPTL 73
Db 1 MLLAIDVRNTHVTWGLLSGSKHAKVVOQWRIRTESEVTADELALIDGLIGDS---ER 57
QY 74 VEGVISSVVPNIMYSLEHMIRKYPKINP-LVVGPGIKTGINKYDNPKEVGADRVNAV 132
Db 58 LAGAALSTVPSVLEHVRIMLDQYWPSPVHLIEPGVGTGIPLLVDNPKVEVGADRVNCL 117
QY 133 AAHEIYKRSLLIIDFTATTTCFAVRENGDYLGAICPGIKVSSEALFEKAALPRVELIK 192
Db 118 AAFHFGQAIVVDFGSSICVDVVSAGFELGAIAPGVQVSDAAARSALRRVELAR 177
QY 193 PAYACKNTISSIQSGIVRYRQVKYLFKELKENLPD-----GRRTTSLVATGGLAK 247
Db 178 PRSVGKNTVECMQAGAVFGFAGLVGLVGRMRQDVEESGDLGNRV---AVVATGHTAP 234
QY 248 LI 249
Db 235 LL 236

RESULT 10

E75516
conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C;Accession: E75516
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: E75516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <WHI>
A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0461
A;Map position: 1
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 26.8%; Score 337; DB 2; Length 262;
Best Local Similarity 33.3%; Pred. No. 2.6e-21;
Matches 80; Conservative 49; Mismatches 89; Indels 22; Gaps 6;

QY 19 ILVDVGNNTNIVLGIYNDT-KLTAEWRLSTDLVRSADYGIQVMNLFQODKLDPTLV--- 74
Db 6 LLAVDIGNTTTVGLADASGALHTWRTNRMLPDDLALQLHGLF-----TLGAP 58
QY 75 --EGVISSVVPNIMYSLEHMIRKYPKINP-LVVGPGIKTGINKYDNPKEVGADRVNAV 132
Db 59 IPRAAVISSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELTDPGVSAGDRCLNF 118
QY 133 AAHEIYKRSLLIIDFTATTTCFAVRENGDYLGAICPGIKVSSEALFEKAALPRVE 189
Db 119 GA-EKYLGLDVAVVDFGTSTNDFVVGRRFLFATGQVSADALFAARAKLPRT 177
QY 190 LIKPAYACKNTISSIQSGIVRYRQVKYLFKELKENLPDGRRTTSVLVATGGLAKLI 249
Db 178 LQAPETAICKNTVHALQSLVFGYAEWVDGLLRIRRAELPG-----EAVAVATGGSRTV 232

RESULT 11

D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic-
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.

Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 26.5%; Score 333.5; DB 2; Length 246;
Best Local Similarity 39.1%; Pred. No. 4.8e-21;
Matches 95; Conservative 39; Mismatches 72; Indels 37; Gaps 11;

QY 20 LVLDVGNNTNIVLGIYNDTKLTAEWRLSTDLVRSADYGIQVMNLFQODKLDPTL----- 73
Db 3 LLVDVGNTHSVFSTEDGKTFRRRLSTGVFTQDE-----LFSH--LHPLLGDMRE 53
QY 74 VEGVISSVVPNIMYSLEHMIRKYPKINP-LVVGPGIKTGINKYDNPKEVGADRVNAV 130
Db 74 VEGVISSVVPNIMYSLEHMIRKYPKINP-LVVGPGIKTGINKYDNPKEVGADRVNAV 130

Db 54 IKGIGVASVPTQNTVIERFSQKYFHISPIWV--KAKNGCVKNVVK--NPSEVGADRVAN 109
QY 131 AVAAHEIYKRSLLIIDFGTATTCFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVEL 190
Db 110 VVAFVKEGKNGLIIDGTAITVDLV-VNGYEGAILPGFFMVHSLFRGTAKLPLVE- 167
QY 191 IKPA-YAICKNTSISSQGIYVYRLQVKYLFKEKLKE---NLDPGRTRTSLVLATGSLA 246
Db 168 VKPADVGVGKDTTEENIRLVGVNGSVYALEGIIGRIKEYVGDLP-----VVLTGQS 218
QY 247 KLI 249
Db 219 KIV 221

RESULT 12

D71326

conserved hypothetical protein TP0431 - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000

C:Accession: D71326

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: D71326

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <COL>

A:Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0431

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 22.5%; Score 282.5; DB 2; Length 273;
Best Local Similarity 29.0%; Pred. No. 1.2e-16;
Matches 69; Conservative 60; Mismatches 98; Indels 11; Gaps 4;

QY 19 ILVDVGNVTNIVLGYNTK-----LTAEWRLSTDVLSADEYGIQVMNLFQDKLDPYLV 74
Db 1 MLLIDVGNHVVFGIQGNGRVCVRELFRLAPDARKTQDBYSLLIHALCERAGVGRASL 60
QY 75 EGVIISSVVPNIMYSLEHMIRKYNPLVVGPGIKTGINIKYDNP--KEVGADRIYNAV 132
61 RDAFISSVVPVLTFTIADAVAQISGVQPVVFGPWAYEHLPRIPVRAEIGTDLVANAV 120
QY 133 AAHEIYKRSLLIIDFGTATTCFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIK 192
Db 121 AAYVHFSACVVVDGTAITFAVDGTLGIQGVAIAPGLRTAVQSLHTGTALPLVPLAL 180
QY 193 PAYAICKNTSISSQGIYVYRLQVKYLFKEKLKENLPDGRTRTSLVLATGSLAKLIN 250
Db 181 PDSVLGKDTTHAVAGVVRGTLFTRAMIAQCQKEL-----GCRCAAVI-TGGLSLRFS 233

RESULT 13

F70165

conserved hypothetical protein BB0527 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999

C:Accession: F70165

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: F70165

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-262 <KLE>
A:Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g268
A:Experimental source: strain B31

Query Match 20.8%; Score 262; DB 2; Length 262;
Best Local Similarity 30.1%; Pred. No. 6.7e-15;
Matches 72; Conservative 54; Mismatches 87; Indels 26; Gaps 9;

QY 20 LVLDVGNVTNIVLGYNTKLTAEWRSLSTDVLSADEYGIQVMNLFQDKLDPYLVGVII 79
Db 9 LIIDIGNTSIAFALFKDQVNLFIKMTNMLRLRYDE----VYSFEEN-FDFN-VNKVFI 62
QY 80 SSVVPNIMYSLEHMIRKYNPLVVGPGIKTGINIKYD---NPKE-----VGADRIYNA 131
Db 63 SSVVPILNETFKNVIFSEFKIKPLFI-----GPDNLVDLTFNPKSKDKLLGSDVFANL 116
QY 132 VAAHEIYK-RSLIIDFGTATTCFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVEL 190
Db 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGLGGINSGPLINFSLLDNAYLIKKEPPI 176
QY 191 IKPAYAICKNTSISSQGIYVYRLQVKYLFKEKLKENLPDGRTRTSLVLATGSLAKLI 249
Db 177 STPNLLERTTSGSVNSGLFYQY----KYLIEGVYRIDIKQYKKFNLIIT-TGGNADLI 230

RESULT 14

B81009

BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: B81009

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qia, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigian, V.; Pizzi, M.
Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: B81009

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-592 <TGT>

A:Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB2075

Query Match 13.4%; Score 168; DB 2; Length 592;
Best Local Similarity 24.7%; Pred. No. 2.1e-06;
Matches 60; Conservative 42; Mismatches 97; Indels 44; Gaps 8;

QY 20 LVLDVGNVTNI-----VLGI--YND-TKLTAEWRLSTDVLSADEYGIQVMNLFQ 65
Db 341 LLLDGGNSRLKAWVWENGTFATVGSAPYRDLSPGLAEWAEKAD--GNRVIGCAVCGBEK 398
QY 66 QDKLDPTLVEGVIISSVVPNIMYSLEHMIRKYNPLVVGPGIKTGINIKYDNPKEVGA 125
Db 399 KAQVOEQARKI---EWLPSSAQAQ-----GTRNHYRHPPEHGS 434
QY 126 DRIVNAVAHEIYKRSLLIIDFGTATTCFCAVRENGDYLGGAICPGIKVSSEALFEKAAL 185
Db 435 DRWFNALGSRFRSRNACVVSGTAVTVDAITDGDGHLGGTITMPGFHLMKESLAVRTANL 494
QY 136 PRVELIKPAYAICKNTSISSQGIYVYRLQVKYLFKEKLKENLPDGRTRTSLVLATGSL 245
Db 495 NR--HAGKRYPPPTTGNNAVASGMDVCGSVMMHGRLEKERTGAG--PVDVIITGGGA 550
QY 246 AKL 248
Db 551 AKV 553

H82031
Probable biotin-[acetyl-CoA-carboxylase] ligase [EC 6.3.4.15] - Neisseria meningitidis
N.Contains: biotin-[acetyl-CoA-carboxylase] ligase [EC 6.3.4.15]
C.Species: Neisseria meningitidis
C.Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C.Accession: H82031
R.Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
S.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A.Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A.Reference number: AB1775; UID:20222556; PMID:10761919
A.Accession: H82031
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-592 <PAR>
A.Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737911
A.Experimental source: serogroup A, strain Z2491
C.Genetics:
A.Gene: birA; NMA0357
Keywords: ligase

Query Match	13.4%;	Score 168;	DB 2;	Length 592;
Best Local Similarity	24.7%;	Pred. No. 2.le-06;		
Matches 60;	Conservative 42;	Mismatches 97;	Indels 44;	Gaps 8;
20	LVLVDVGNNTI-----VLGI--YND-TKLTAEWRLSTDVLRSADEYGIQVMNLFQ	65		
	:	:	:	:
341	LLLDGGSRLKNAWENGTFATVGSAPYRDLSPGAEWAQVD--GNVRIVGCAVCGEFK	398		
	:	:	:	:
66	ODKLDPTLVEGVLISSVVENIMKSYLEHMIRKYFLNPLVPGTGKTKINKYDNPKEVGA	125		
	:	:	:	:
399	KAQVQEQLARKI---EWLPSSAQAAL-----GIRNHYRHPPEHGS	434		
	:	:	:	:
126	DRIVNVAAREHYIKRSLIIDFGTATTFCAVRENGDYLGGGAICPGIKVSSSEALFEKAAL	185		
	:	:	:	:
435	DRWFNALGSRFRSRNACVVCVSGTAVTDALTDGCHYLGGTIMPGFHLMKESLAVRTANL	494		
	:	:	:	:
186	PRVELIKPAYAICKNTYISSIQSGIYRYLRQVKYLFKELKNLPDGRTRTSLVATGGGL	245		
	:	:	:	:
495	NR--HAGKRYPPFTTTGNVAAGMDAVCGSVMMHMLKKEKTGAGK--PVDVIITGGGA	550		
	:	:	:	:
246	AKL 248			
551	AKV 553			
Db				

Search completed: June 24, 2003, 22:02:57
Job time : 10.1185 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 25.667 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-3
Perfect score: 1257
Sequence: 1 NKRAAFMLLLFLRSVLKVL.....GRRTRSLVLATGSLAKLIN 250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1257	100.0	250	AAU91150	Clostridium acetob
2	695	55.3	256	AAU91175	Pantothenate kinas
3	665.5	52.9	262	AAU91170	Pantothenate kinas
4	648.5	51.6	254	AAU91171	Pantothenate kinas
5	646.5	51.4	258	AAU01243	B. subtilis novel
6	646.5	51.4	258	AAU91149	Bacillus subtilis
7	631.5	50.2	255	AAU91154	Geobacter sulfurre
8	630.5	50.2	259	ABBA7661	Listeria monocytog
9	629.5	50.1	258	AAU91172	Pantothenate kinas
10	623	49.6	233	AAU91163	Pantothenate kinas

11	513	40.8	219	23	AAU91176	Pantothenate kinas
12	498.5	39.7	265	23	AAU91151	Streptomyces coeli
13	460.5	36.6	260	23	AAU91173	Pantothenate kinas
14	428.5	34.1	258	23	AAU91153	Rhodobacter capsul
15	392	31.2	272	22	AAU91152	Mycobacterium tube
16	389	30.9	272	23	AAU91152	Mycobacterium tube
17	337	26.8	262	23	AAU91155	Deinococcus radiop
18	333.5	26.5	246	23	AAU91156	Thermotoga maritim
19	327.5	26.1	212	23	AAU91177	Pantothenate kinas
20	282.5	22.5	273	23	AAU91157	Treponema pallidum
21	262	20.8	262	23	AAU91158	Borrelia burgdorfe
22	229.5	18.3	257	23	AAU91174	Pantothenate kinas
23	169	13.4	241	23	AAU91179	Pantothenate kinas
24	169	13.4	455	20	AAU91167	Neisseria gonorrhoe
25	169	13.4	455	21	AAU91167	Neisseria gonorrhoe
26	169	13.4	460	23	AAU91167	Pantothenate kinas
27	169	13.4	592	20	AAU91167	Neisseria gonorrhoe
28	169	13.4	592	21	AAU91167	Neisseria gonorrhoe
29	168	13.4	455	21	AAU91167	Neisseria meningit
30	168	13.4	592	20	AAU91167	Neisseria meningit
31	168	13.4	592	20	AAU91167	Neisseria meningit
32	168	13.4	592	21	AAU91167	Neisseria meningit
33	168	13.4	592	21	AAU91167	Neisseria meningit
34	168	13.4	592	23	AAU91166	Pantothenate kinas
35	168	13.4	592	23	AAU91169	Pantothenate kinas
36	167.5	13.3	244	23	AAU91168	Pantothenate kinas
37	167.5	13.3	389	21	AAU91168	Neisseria meningit
38	159.5	12.7	248	23	AAU91164	Pantothenate kinas
39	156	12.4	249	23	AAU91178	Pantothenate kinas
40	141	11.2	267	23	AAU91162	Bordetella pertussis
41	136.5	10.9	249	23	AAU91182	Pantothenate kinas
42	130.5	10.4	189	20	AAU91182	Neisseria meningit
43	123.5	9.8	229	23	AAU91159	Aquifex aeolicus p
44	121.5	9.7	257	23	AAU91160	Synechocystis pant
45	108.5	8.6	242	23	AAU91180	Pantothenate kinas

ALIGNMENTS

RESULT 1
AAU91150
ID AAU91150 standard; Protein; 250 AA.

XX AAU91150;

XX AC AAU91150;

XX DT 05-JUN-2002 (first entry)

XX DE Clostridium acetobutylicum pantothenate kinase Coax.

XX DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX OS Clostridium acetobutylicum.

XX PN WO200216601-A2.

XX PD 28-FEE-2002.

XX XX 24-AUG-2001; 2001WO-US26531.

XX XX 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX DR WPI; 2002-269358/31.

XX XX Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 68-69; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 250 AA;

RESULT 2	
AAU911175	AAU911175 standard; Protein; 256 AA.
XX	
XX	AAU911175;
XX	
XX	05-JUN-2002 (first entry)
XX	
XX	Pantothenate kinase (Coax) #13.
XX	
KW	Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX	
OS	Clostridium difficile.
XX	
PN	WO200216601-A2.
XX	
PD	28-FEB-2002.
XX	
PF	24-AUG-2001; 2001WO-US26531.
XX	
PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WPI; 2002-269358/31.
DR	N-PSDB; ABK54196.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprises

CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.

XX Sequence 255 AA;

SQ Query Match 50.2%; Score 631.5; DB 23; Length 255;
Best Local Similarity 54.7%; Pred. No. 2.8e-65;
Matches 127; Conservative 41; Mismatches 59; Indels 5; Gaps 1;

QY 18 VILVLDVGNNTNIVLGIYNDTKLTAEWRLSTDVLSADEYGVQVNLFOODKLDPTLVGV 77
DB 1 MLLVIDVGNNTNIVLGIYDGERLVRDWRVSTDKARTDEYGLINELFRLAGLGLDQIRAV 60
QY 78 IISVVPNIMYSLEHMIRKYEKINPLVVGPGIKTGINIKYDNPKREVAGADRIVNAVAHEI 137
DB 61 IISVVPPLTGVLERLSLGYGMRPLVVGPGIKTGMPLOYDNPREVGADRIVNAVAGYEK 120
QY 138 YKRSIIIDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 197
DB 121 YRTSLIIVDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 180
QY 198 CKNTISSQSGIVVYRLQVKYLPFKLENLPDGRRTTSLVLTATGGGLAKLI 249
DB 181 ARNTVSMQAGIYGYVGLVDEIVTRMKAESKDAPR-----VIATGGGLASLI 227

RESULT 8

ID ABBA7661 standard; Protein; 259 AA.

XX ABBA7661;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #365.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;

XX WPI; 2002-010914/01.

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides

XX Claim 6; SEQ ID No 366; 192pp; French.

XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGB-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 259 AA;

Query Match 50.2%; Score 630.5; DB 23; Length 259;
Best Local Similarity 51.3%; Pred. No. 3.8e-65;
Matches 119; Conservative 54; Mismatches 54; Indels 5; Gaps 1;

QY 18 VILVLDVGNNTNIVLGIYNDTKLTAEWRLSTDVLSADEYGVQVNLFOODKLDPTLVGV 77
DB 1 MLLVIDVGNNTNIVLGIYDGERLVRDWRVSTDKARTDEYGLINELFRLAGLGLDQIRAV 60
QY 78 IISVVPNIMYSLEHMIRKYEKINPLVVGPGIKTGINIKYDNPKREVAGADRIVNAVAHEI 137
DB 61 IISVVPPIHMETMCMVRYENIRPLIVGPGIKTGLNKLVDNPREIGSDRIVNAASEE 120
QY 138 YKRSIIIDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 197
DB 121 YGTPVIVVDFGTATTCFAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 180
QY 198 CKNTISSQSGIVVYRLQVKYLPFKLENLPDGRRTTSLVLTATGGGLAKLI 249
DB 181 GKSTVSMQAGIYGYVGLVDEIVTRMKAESKDAPR-----QSNASPVVVGGLARMI 227

RESULT 9

AAU91172
ID AAU91172 standard; Protein; 258 AA.

XX AAU91172;

DT 05-JUN-2002 (first entry)

DE Pantothenate kinase (Coax) #10.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Bacillus stearothermophilus.

OS WO200216601-A2.

PN 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

PR 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

DR N-PSDB; ABK54193.

XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein
XX
XX
PS Claim 10; Page 101-102; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 258 AA;
SQ
Query Match 50.1%; Score 629.5; DB 23; Length 258;
Best Local Similarity 53.8%; Pred. No. 4.9e-65;
Matches 126; Conservative 40; Mismatches 59; Indels 9; Gaps 2;
QY 18 VILVLDVGNINVLGYNDTKLFAEWRSLDVLRSADYGIQVNNLFQODKLPDPTLVEGV 77
DB 1 MIFVLDVGNINVLGYNDTKLFAEWRSLDVLRSADYGIQVNNLFQODKLPDPTLVEGV 60
QY 78 IISVVNPIMYSLEHMIRKVFKNPLVGVGPKTGKINIKYDNPKEVGADRIVNAAHEI 137
DB 61 IISVVNPIMFALERCKYFHIEPQIVGPMKGTGKINIKYDNPKEVGADRIVNAAHEI 120
QY 138 YKRSLLIIDFGTATTCFAVRENGDYLGGATCGIKVSSSEALFEKAALPRVELIKPAYAI 197
DB 121 YGSLIIVDFGTATTTCYINEHQYMGGAIPGIMISTEALFARAALPRIETARDDII 180
QY 198 CKNTISSIOGIVRYLRQVKLYFE--KLKENIPDGRRTSIVLATGGLAKLI 249
DB 181 GKNTVSAMQAGILYGVQVEGIVSRMAKSKIP-----PKVIATGGLAPLI 227
RESULT 10
AAU91163
ID AAU91163 standard; Protein; 233 AA.
XX
XX AAU91163;
XX
XX 05-JUN-2002 (first entry)
XX
XX Pantothenate kinase (Coax) #1.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX Bacillus subtilis.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
XX
XX 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX Yocum RR, Patterson TA;
XX
XX WPI; 2002-269358/31.
XX
XX N-PSDB; ABK54169.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein
XX
XX
PS Disclosure; Page 81-82; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 233 AA;
SQ
Query Match 49.6%; Score 623; DB 23; Length 233;
Best Local Similarity 55.8%; Pred. No. 2.4e-64;
Matches 116; Conservative 44; Mismatches 48; Indels 0; Gaps 0;
QY 18 VILVLDVGNINVLGYNDTKLFAEWRSLDVLRSADYGIQVNNLFQODKLPDPTLVEGV 77
DB 1 MLIVDVGNTNTVLGYHDGKLEYHRIETSRHKTDEFGMILRSLFDHSGLMFEQIDGI 60
QY 78 IISVVNPIMYSLEHMIRKVFKNPLVGVGPKTGKINIKYDNPKEVGADRIVNAAHEI 137
DB 61 IISVVNPIMFALERCKYFHIEPQIVGPMKGTGKINIKYDNPKEVGADRIVNAAHEI 120
QY 138 YKRSLLIIDFGTATTCFAVRENGDYLGGATCGIKVSSSEALFEKAALPRVELIKPAYAI 197
DB 121 YGNPLIWDFTGATTTCYIDENQYMGGAIPGITTISTEALYSRAAKLPRIETRPDNI 180
QY 198 CKNTISSIOGIVRYLRQVKLYFEKLK 225
DB 181 GKNTVSAMQAGILYGVQVEGIVSRMAK 208
RESULT 11
AAU91176
ID AAU91176 standard; Protein; 219 AA.
XX
XX AAU91176;
XX
XX 05-JUN-2002 (first entry)
XX
XX Pantothenate kinase (Coax) #14.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX Dehalococcoides ethenogenes.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
XX
XX 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX Yocum RR, Patterson TA;
XX
XX WPI; 2002-269358/31.
XX
XX N-PSDB; ABK54197.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein

XX Claim 10; Page 106-107; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

SQ Sequence 219 AA;

Query Match 40.8%; Score 513; DB 23; Length 219;

Best Local Similarity 45.6%; Pred. No. 1.8e-51;

Matches 99; Conservative 47; Mismatches 69; Indels 2; Gaps 1;

QY 19 ILVLDVGNNTNVLGIYNDTKLTAEWRLSTDVLRSADEYGVGMNLFQDKLDPVLVEGVI 78

DB 5 LVADIGNTSVNIIGIFEGEKLALNHLGSAVRAQWADYASLLGLLQHAGHPLELNKVI 64

QY 79 ISSVVPNTMYSLEHMIRKRYKINPLVWPGIKTGINKYDNPKEVGADRIYVNAVAHEIY 138

DB 65 MCSVVPPLTTTFEEVFKSYKKAAPLVVGAGIKSGVKVRMDNPREVGADRIYVNAARVLY 124

QY 139 KRSLLIIDFGTATTCFAVRENGDYLGGACPGIKVSSEALFEKAALPRVELIKPAYAIC 198

DB 125 PGACIIVDMGTATFTDLSEGGAYIGGAIPGAIATSAQAIAEKTSKLPKBIIRPAKVI 184

QY 199 KNTISSQSGIVRYLROVKYLFKEKLKLENLPDGRRT 235

DB 185 SNTVSQSGIVFYIGVIGVELVRRIQTEL--GQKTR 219

RESULT 12

AAU91151

ID AAU91151 standard; Protein; 265 AA.

AC AAU91151;

XX 05-JUN-2002 (first entry)

XX Streptomyces coelicolor pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Streptomyces coelicolor.

OS WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises

XX contacting composition comprising pantothenate kinase (Coax) protein

XX with test compound and identifying inhibitor of the Coax protein -

PS Claim 10; Page 69-70; 128pp; English.

CC The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX Sequence 265 AA;

Query Match 39.7%; Score 498.5; DB 23; Length 265;

Best Local Similarity 43.3%; Pred. No. 1.2e-49;

Matches 104; Conservative 50; Mismatches 73; Indels 13; Gaps 4;

QY 18 VILVDGNTNVLGIYNDTKLTAEWRLSTDVLRSADEYGVGMNLFQDKLDPVL---- 73

DB 1 MLLTIDVGNTHVLGLFDGEDIVHWRISTDSRRTADELAVLQGLM---GMHPLLGDEL 57

QY 74 ---VEGVTISSVVPNTMYSLEHMIRKRYKINPLV--VPGIKTGINKYDNPKEVGADRIY 129

DB 58 GDGIDGIAICATVPSPVLHREVTTRYGDPVAVLVEPGVKTPILTDPKPEVGADRII 117

QY 130 NAVAHAHEYKRSLLIIDFGTATTCFAVRENGDYLGGACPGIKVSSEALFEKAALPRVE 189

DB 118 NAVAHAELYGGAIPVDEGTATTFDAVSARGEYIGGVIAPIEISVEALGVKGAOLRIE 177

QY 190 LIKPAYATCKNTISSQSGIVRYLROVKYLFKEKLKLENLPDGRRTTSLVLATGGLAKLI 249

DB 178 VARPSRVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDDPDVT--VIATGGLAPMV 235

RESULT 13

AAU91173

ID AAU91173 standard; Protein; 260 AA.

AC AAU91173;

XX 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #11.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Caulobacter crescentus.

OS WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX N-PSDB; ABK54194.

XX Identifying potential antibiotic or antimicrobial agent, comprises

XX contacting composition comprising pantothenate kinase (Coax) protein

XX with test compound and identifying inhibitor of the Coax protein -

PS Claim 10; Page 102-103; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.

XX Sequence 260 AA;

Query Match 36.6%; Score 460.5; DB 23; Length 260;

Best Local Similarity 39.5%; Pred. No. 3.4e-45;

Matches 92; Conservative 53; Mismatches 81; Indels 7; Gaps 2;

QY 18 VILVLDVGNINVLGYNDTKLTAEWRLSTDLVRSADYGIQVNNLFQDDKLDPTLVEGV 77
DB 1 MLIAICGGNTNTFAIHGASVQAQRSAFESTRTADEYVWLSQGLGFRADAV 60
QY 78 IISVVPNIMYSLEHMIRYFKINPLVVGPGIKTGINKYDNPKEVGADRIVNAVAHEI 137
DB 61 IISVVVPQSIFNLRLNRRYFNVEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAICAAV 120
QY 138 YKRSLLIIDFGTATTCVAVRENGDYLGGAIKPCIKVSSSEALFEKAALPRVELIKPA--Y 195
DB 121 YPGPLVVIDSGTATTFDIAADGAFEGGIAPGINLSMQALHEAAKLPRIATQRPAGNR 180
QY 196 AICKNTISSIOSGIVRYLRQVKYLFKLENLPDGRRTSLVLATGLAKL 248
DB 181 IVGTDRVSAMQSGVFWGYISLIEGLVARIK-----AERGPMTVIATGGVASL 228

RESULT 14

AAU91153
ID AAU91153 standard; Protein; 258 AA.

XX AC AAU91153;

XX DT 05-JUN-2002 (first entry)

XX DE Rhodobacter capsulatus pantothenate kinase Coax.

XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX OS Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX PN Rhodobacter capsulatus.

XX PP WO200216601-A2.

XX XX 28-FEB-2002.

XX XX 24-AUG-2001; 2001WO-US26531.

XX XX 24-AUG-2000; 2000US-227860P.

XX XX 20-MAR-2001; 2001US-0813453.

XX XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX XX Yocum RR, Patterson TA;

XX XX WPI; 2002-269358/31.

XX DR Identifying potential antibiotic or antimicrobial agent, comprises

XX PT contacting composition comprising pantothenate kinase (Coax) protein

XX PT with test compound and identifying inhibitor of the Coax protein -

XX XX Claim 10; Page 71-72; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic

XX CC comprising contacting an assay composition comprising a pantothenate

XX CC kinase (Coax) protein with a test compound, and determining the ability

XX CC of the test compound to inhibit the activity of the Coax protein, an

XX CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.

XX Sequence 258 AA;

Query Match 34.1%; Score 428.5; DB 23; Length 258;

Best Local Similarity 39.7%; Pred. No. 1.9e-41;

Matches 93; Conservative 45; Mismatches 89; Indels 7; Gaps 3;

QY 18 VILVLDVGNINVLGYNDTKLTAEWRLSTDLVRSADYGIQVNNLFQDDKLDPTLVEGV 77
DB 1 MLLCIDCGNTNTVFSVWDGDTFAATWRIATDHERTADEYFVWLTMLQKLGLOGRISE-A 59
QY 78 IISVVPNIMYSLEHMIRYFKINPLVVG-PGIKTGINKYDNPKEVGADRIVNAVAHE 136
DB 60 IISSTAPRVVFNLRVLCNRYFDCRPVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYD 119
QY 137 IVKRSLLIIDFGTATTCVAVRENGDYLGGAIKPCIKVSSSEALFEKAALPRVELIKPAYA 196
DB 120 RHGGDLIVVDFTGATTFDVPDCAVIGGVIAPCVNLSEALHMAAALPHVDVTKPGV 179
QY 197 ICKNTISSIOSGIVRYLRQVKYLFKLENLPDGRRTSLVLATGLAKLIN 250
DB 180 IGNTVACIOSGVWGYIGLVGIVRQIRM-----ERDRPMKVIAATGGLASLFD 228

RESULT 15

AA81225
ID AA81225 standard; Protein; 272 AA.

XX AC AA81225;

XX DT 04-SEP-2001 (first entry)

XX DE Mycobacterium tuberculosis potential drug target protein SEQ ID 276.

XX KW Drug target; growth; organism viability; characterisation.

XX OS Mycobacterium tuberculosis.

XX PN WO200135317-A1.

XX XX 17-MAY-2001.

XX XX 13-NOV-2000; 2000WO-US31152.

XX XX 12-NOV-1999; 99US-0165086.

XX XX 12-NOV-1999; 99US-0165124.

XX XX 01-FEB-2000; 2000US-0179531.

XX XX (REGC) UNIV CALIFORNIA.

XX XX Eisenberg D, Rotstein SH, Marcotte EM;

XX XX WPI; 2001-329193/34.

XX XX N-PSDB; AAH52076.

XX PT Identifying nucleotide or polypeptide sequence for use as drug target,

XX PT involves providing algorithm that analyzes a functional relationship

XX PT between nucleotide or polypeptide sequences, and comparing the

XX PT sequences -

XX PS Disclosure; Page 188; 207pp; English.

XX CC This invention relates to a method for identifying a nucleotide or

XX CC polypeptide sequence that may be a drug target, or essential for growth

XX CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52052

XX CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium

XX CC tuberculosis proteins which are potential drug targets. The DNA and

XX CC protein sequences are used to illustrate the method of the invention. The

CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.

xx
cc
SQ Sequence 272 AA;

Query Match 31.2%; Score 392; DB 22; Length 272;
Best Local Similarity 34.6%; Pred. No. 4e-37;
Matches 82; Conservative 62; Mismatches 85; Indels 8; Gaps 3;
QY 18 VILVDVGNINVLGIYN---DFKLTAEWRLSTDVLSADEYGIQVMNLFQODKLDPTL 73
DB 1 VLLAIDVRNTHVTVGLLSGMEHAKVQVQWRIRTESEVTADLALTDGLIGDS---ER 57
QY 74 VEGVLISSVWPNIIMYSLEHMIRKYFKINP-LVVGPGIKTGINKYDNPKEVGADRIVNAV 132
DB 58 LTGTAALSTVPSVLHEVRIMLDQVWPSPVHVLIEPGVRTGIPLLVDNPKVVGADRIVNCL 117
QY 133 AAHEIYKESLIIIDFGTATTCVAVRENGDYLGGAICPGIKVSSSEALFEKAALPRVELIK 192
DB 118 AAYDRFRKAAIVDFGSSICVDVVSAAKGEFLGGAIFAGVQVSSDAAAARSAAALRRVELAR 177
QY 193 PAYAICKNTISSIOSGIVRYRLQVKYLFKELKENLPDGRTRTSLVLATGGLAKLI 249
DB 178 PRSVVGKNTVECMQAGAVFGFAGLVGLVGRREDVSGFSDVDHDAIVATGHTAPLL 234

Search completed: June 24, 2003, 21:46:11
Job time : 26.667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.1803 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-7
Perfect score: 1279
Sequence: 1 MLLVIVGNTNIVLGIYDGE.....AAVEYLTLEGLRILYERNRE 255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1279	100.0	255	23 AAU91154	Geobacter sulfurre
2	834	65.2	258	23 AAU91172	Pantothenate kinas
3	785.5	61.4	256	23 AAU91175	Pantothenate kinas
4	783	61.2	262	23 AAU91170	Pantothenate kinas
5	771	60.3	254	23 AAU91171	Pantothenate kinas
6	756	59.1	258	22 AAU01243	B. subtilis novel
7	756	59.1	258	23 AAU91149	Bacillus subtilis
8	749	58.6	259	23 ABB47661	Listeria monocytog
9	650	50.8	233	23 AAU91163	Pantothenate kinas
10	631.5	49.4	250	23 AAU91150	Clostridium acetob

11	629	49.2	265	23 AAU91151	Streptomyces coeli
12	561	43.9	260	23 AAU91173	Pantothenate kinas
13	558	43.6	219	23 AAU91176	Pantothenate kinas
14	554	43.3	258	23 AAU91153	Rhodobacter capsul
15	509.5	39.8	272	22 AAU91152	Mycobacterium tube
16	505.5	39.5	272	22 AAG81225	Mycobacterium tube
17	414.5	32.4	262	23 AAU91155	Deinococcus radiop
18	371.5	29.0	273	23 AAU91157	Treponema pallidum
19	361.5	28.3	246	23 AAU91156	Thermotoga maritim
20	338.5	26.5	212	23 AAU91177	Pantothenate kinas
21	301	23.5	257	23 AAU91174	Pantothenate kinas
22	266.5	20.8	262	23 AAU91158	Borrelia burgdorfe
23	208	16.3	244	23 AAU91168	Pantothenate kinas
24	172.5	13.5	241	23 AAU91179	Pantothenate kinas
25	163.5	12.8	249	23 AAU91182	Pantothenate kinas
26	159	12.4	249	23 AAU91178	Pantothenate kinas
27	148	11.6	267	23 AAU91162	Bordetella pertussis
28	143	11.2	455	20 AAY38617	Neisseria gonorrhoe
29	143	11.2	455	21 AAY74908	Neisseria gonorrhoe
30	143	11.2	460	23 AAU91167	Pantothenate kinas
31	143	11.2	592	20 AAY38618	Neisseria gonorrhoe
32	143	11.2	592	21 AAY74911	Neisseria gonorrhoe
33	141	11.0	248	23 AAU91164	Pantothenate kinas
34	141	11.0	455	21 AAY74910	Neisseria meningit
35	141	11.0	592	20 AAY38616	Neisseria meningit
36	141	11.0	592	21 AAY74913	Neisseria meningit
37	141	11.0	592	23 AAU91166	Pantothenate kinas
38	139.5	10.9	229	23 AAU91159	Aquifex aeolicus p
39	139.5	10.9	389	21 AAY74909	Neisseria meningit
40	139.5	10.9	592	20 AAY38615	Neisseria meningit
41	139.5	10.9	592	21 AAY74912	Neisseria meningit
42	139.5	10.9	592	23 AAU91169	Pantothenate kinas
43	128	10.0	242	23 AAU91180	Pantothenate kinas
44	126	9.9	257	23 AAU91160	Synechocystis pant
45	114.5	9.0	189	20 AAY38614	Neisseria meningit

ALIGNMENTS

RESULT 1
AAU91154
ID AAU91154 standard; Protein; 255 AA.
AC AAU91154;
XX
XX
XX
DT 05-JUN-2002 (first entry)
XX
DE Geobacter sulfurreducens pantothenate kinase Coax.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Geobacter sulfurreducens.
PN
PN WO200216601-A2.
PD
PD 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PA Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -

XX

PS Claim 10; Page 72-73; 128pp; English.

XX

CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.

XX

SQ Sequence 255 AA;

Query Match 100.0%; Score 1279; DB 23; Length 255;
 Best Local Similarity 100.0%; Pred. No. 4.2e-126;
 Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
 |||||

1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
 |||||

61 IISVVPPPLTGVLERLSLGYFGMRPLVVGPIQYDNPREGVADRIYNVAVGYEK 120
 |||||

61 IISVVPPPLTGVLERLSLGYFGMRPLVVGPIQYDNPREGVADRIYNVAVGYEK 120
 |||||

121 YRTSLIIVDFGATTDDYNNRGEYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAII 180
 |||||

121 YRTSLIIVDFGATTDDYNNRGEYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAII 180
 |||||

181 ARTVNSMQAGIYGVVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIEAVEEY 240
 |||||

181 ARTVNSMQAGIYGVVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIEAVEEY 240
 |||||

241 LLEGRLILYERNRE 255
 |||||

241 LLEGRLILYERNRE 255
 |||||

RESULT 2

AAU91172
 ID AAU91172 standard; Protein; 258 AA.

XX AAU91172;

XX AC AAU91172;

XX DT 05-JUN-2002 (first entry)

Pantothenate kinase (Coax) #10.

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Bacillus stearothermophilus.

XX WO200216601-A2.

XX PN WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX XX 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX XX Yocum RR, Patterson TA;

XX PI WPI; 2002-269358/31.

XX DR N-PSDB; ABK54193.

XX XX Identifying potential antibiotic or antimicrobial agent, comprises

XX DR N-PSDB; ABK54196.

XX PT

PT

PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -

XX

PS Claim 10; Page 101-102; 128pp; English.

XX

CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.

XX

SQ Sequence 258 AA;

Query Match 65.2%; Score 834; DB 23; Length 258;
 Best Local Similarity 62.7%; Pred. No. 2.8e-79;
 Matches 160; Conservative 40; Mismatches 55; Indels 0; Gaps 0;

QY 1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
 ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 MIFVLDVGNNTVLGYDGDGLKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
 ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 IISVVPPPLTGVLERLSLGYFGMRPLVVGPIQYDNPREGVADRIYNVAVGYEK 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 61 IISVVPPPLTGVLERLSLGYFGMRPLVVGPIQYDNPREGVADRIYNVAVGYEK 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 YRTSLIIVDFGATTDDYNNRGEYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAII 180
 ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 121 YGSPLIIVDFGATTVCYINEHKQYMGGAIPGIMISTEALFARAAKLPRIETARPDII 180
 ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 ARTVNSMQAGIYGVVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIEAVEEY 240
 :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 181 GNTVSAMQAGILYGVGVGEIVSRMKAESKIPPKVIATGGLAPLIASESDIIDVDPF 240
 ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 241 LTLEGLRILYERNRE 255
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 241 LTLEGLRILYERNRE 255
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3

AAU91175
 ID AAU91175 standard; Protein; 256 AA.

XX AAU91175;

XX AC AAU91175;

XX DT 05-JUN-2002 (first entry)

XX DT Pantothenate kinase (Coax) #13.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.

KW Clostridium difficile.

OS WO200216601-A2.

XX PN WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX XX 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PA Yocum RR, Patterson TA;

XX PI WPI; 2002-269358/31.

XX DR N-PSDB; ABK54196.

XX DR

XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
XX Claim 6; Page 105; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 256 AA;
XX
XX Query Match 61.4%; Score 785.5; DB 23; Length 256;
XX Best Local Similarity 60.9%; Pred. No. 3.5e-74;
XX Matches 156; Conservative 44; Mismatches 55; Indels 1; Gaps 1;
XX
XX QY 1 MLLVVDVGNNTNIVLGIYDGERLVRDVRVSTDKARTDEYGLINELFRLAGLGLDQIRAV 60
XX Db 1 MLLVFDVGNNTNIVLGIYKDKLVNWRKTDREKTSDEYGLISNLFYDYNVISDIDDV 60
XX
XX QY 61 IISVVPPLTGVLERLSLGYFGMRPLVVGPGIKTGMPIQYDNPREGVADRIVNAVAGYEK 120
XX Db 61 IISVVPNMVHLENFCIKYCKQPLVVGPGIKTGLNPKYDNPREGVADRIVNAVAGIEK 120
XX
XX QY 121 YRSLIIVDFGTATTFDYVNRKGEYCGGATAPGLVISTEALFQASKLPRVDIIRPSAII 180
XX Db 121 YGAPSLIVDFGTATTFCAISEKGEYLGTTAPGKISSEALFQASKLPRVELAKPGMTI 180
XX
XX QY 181 ARNTVNSMQAGIYGYVGLVDEIVTRKAE-SKDAPRVIAATGLASLIAPESKTIEAVEE 239
XX Db 181 CKSTVSAMQSGIIYGYVGLVDKIISIMKKELNCDVVKVIATGGGLAKLIASETRKSIDYVDG 240
XX
XX QY 240 YLLEGLRLIYERNRE 255
XX Db 241 FLTLEGLRLIYERNQE 256
XX
XX
XX RESULT 4
XX AAU91170 standard; Protein; 262 AA.
XX AC AAU91170;
XX
XX DT 05-JUN-2002 (first entry)
XX
XX DE Pantothenate kinase (Coax) #8.
XX
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX OS Bacillus anthracis.
XX
XX PN WO200216601-A2.
XX
XX PD 28-FEB-2002.
XX
XX PF 24-AUG-2001; 2001WO-US26531.
XX
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX PI Yocum RR, Patterson TA;
XX

DR WPI; 2002-269358/31.
XX N-PSDB; ABK54191.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
XX Claim 8; Page 98-99; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 262 AA;
XX
XX Query Match 61.2%; Score 783; DB 23; Length 262;
XX Best Local Similarity 57.3%; Pred. No. 6.6e-74;
XX Matches 145; Conservative 51; Mismatches 57; Indels 0; Gaps 0;
XX
XX QY 1 MLLVVDVGNNTNIVLGIYDGERLVRDVRVSTDKARTDEYGLINELFRLAGLGLDQIRAV 60
XX Db 1 MIFVLDVGNNTNIVLGVFEEGELRQHRMETDRHTEDEYGLVYKQLLEHGLSTEDYKGI 60
XX
XX QY 61 IISVVPPLTGVLERLSLGYFGMRPLVVGPGIKTGMPIQYDNPREGVADRIVNAVAGYEK 120
XX Db 61 IVSSVWPPIFALECMCEKFKIKPLVVGPGIKTGLNPKYDNPREGVADRIVNAVAGIHL 120
XX
XX QY 121 YRSLIIVDFGTATTFDYVNRKGEYCGGATAPGLVISTEALFQASKLPRVDIIRPSAII 180
XX Db 121 YGSLIIVDFGTATTFDYVNRKGEYCGGATAPGLVISTEALFQASKLPRVDIIRPSAII 180
XX
XX QY 181 ARNTVNSMQAGIYGYVGLVDEIVTRKAE-SKDAPRVIAATGLASLIAPESKTIEAVEE 240
XX Db 181 GKNIVSAMQSGIIYGYVGLVDEIVTRKAE-SKDAPRVIAATGLASLIAPESKTIEAVEE 240
XX
XX QY 241 LTLEGLRLIYERN 253
XX Db 241 LTLKGLYLYERN 253
XX
XX
XX RESULT 5
XX AAU91171
XX ID AAU91171 standard; Protein; 254 AA.
XX AC AAU91171;
XX
XX DT 05-JUN-2002 (first entry)
XX
XX DE Pantothenate kinase (Coax) #9.
XX
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX OS Bacillus halodurans.
XX
XX PN WO200216601-A2.
XX
XX PD 28-FEB-2002.
XX
XX PF 24-AUG-2001; 2001WO-US26531.
XX
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX PI Yocum RR, Patterson TA;
XX

AAU91163
ID AAU91163 standard; Protein; 233 AA.

XX AC AAU91163;
XX DT 05-JUN-2002 (first entry)
XX DE
XX DE Pantothenate kinase (Coax) #1.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS
XX OS Bacillus subtilis.
XX PN
XX PN WO200216601-A2.
XX PD
XX PD 28-FEB-2002.
XX PF
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI
XX PI Yocum RR, Patterson TA;
XX DR
XX DR WPI; 2002-269358/31.
XX DR N-PSDB; ABK54169.
XX PT
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -
XX PS
XX PS Disclosure; Page 81-82; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.

XX SQ Sequence 233 AA;
Query Match 50.8%; Score 650; DB 23; Length 233;
Best Local Similarity 57.4%; Pred. No. 5.5e-60;
Matches 124; Conservative 38; Mismatches 54; Indels 0; Gaps 0;
QY 1 MLIVDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLQIRAV 60
DB 1 MLIVDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLQIRAV 60
QY 61 IISVVVPLTGVLESLGICFMRPLVVGPGIKTGMPQIYDNPREGADRIYNVAVGEK 120
DB 61 IISVVVPLTGVLESLGICFMRPLVVGPGIKTGMPQIYDNPREGADRIYNVAVGEK 120
QY 121 YRTSLIIVDFGATTFDYVNRKGYCGGAIAPLVISTEALFQRASKLPVRDIIIRPSAI 180
DB 121 YRTSLIIVDFGATTFDYVNRKGYCGGAIAPLVISTEALFQRASKLPVRDIIIRPSAI 180
QY 181 ARNTVNSMQAGIYGVGLVDEIVTRMKAESKDAPR 216
DB 181 ARNTVNSMQAGIYGVGLVDEIVTRMKAESKDAPR 216
GANTVSAMQSGILFGYGVGVEGIVRKMVKWQAKQDP 216

RESULT 10
AAU91150
ID AAU91150 standard; Protein; 250 AA.

XX AC

AAU91150;

05-JUN-2002 (first entry)

Clostridium acetobutylicum pantothenate kinase Coax.

Pantothenate kinase; Coax; antibiotic; antimicrobial;
Pantothenate kinase modulator; coenzyme A; bactericidal compound.

Clostridium acetobutylicum.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

Identifying potential antibiotic or antimicrobial agent, comprises
contacting composition comprising pantothenate kinase (Coax) protein
with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 68-69; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic
comprising contacting an assay composition comprising a pantothenate
kinase (Coax) protein with a test compound, and determining the ability
of the test compound to inhibit the activity of the Coax protein, an
essential enzyme for the production of coenzyme A. Coax protein is a
valuable target for identifying bactericidal compounds. Coax modulating
agents can be used in an infectious animal model to determine the
efficacy, toxicity, or side effects of treatment with such an agent. This
is the amino acid sequence of a pantothenate kinase (Coax) protein
described in the invention.

Sequence 250 AA;

Query Match 49.4%; Score 631.5; DB 23; Length 250;
Best Local Similarity 54.7%; Pred. No. 5.4e-58;
Matches 127; Conservative 41; Mismatches 59; Indels 5; Gaps 1;

QY 1 MLIVDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLQIRAV 60
DB 18 VILVLDVGNNTNVLGIYDNTKLTAEWRLSTDLRSADYGVGMNLFQDKLDPVGEV 77
QY 61 IISVVVPLTGVLESLGICFMRPLVVGPGIKTGMPQIYDNPREGADRIYNVAVGEK 120
DB 78 IISVVVPLTGVLESLGICFMRPLVVGPGIKTGMPQIYDNPREGADRIYNVAVGEK 137
QY 121 YRTSLIIVDFGATTFDYVNRKGYCGGAIAPLVISTEALFQRASKLPVRDIIIRPSAI 180
DB 138 YKRSIIIDFGATTFCAVRRENGDYLGAICFIVSSEALFEKAKLPVELIKPAYAI 197
QY 181 ARNTVNSMQAGIYGVGLVDEIVTRMKAESKDAPR 227
DB 198 CKNTISSQSGIYRYLRQVYLFKLEKLPDGRRTSLVLTATGLAKLI 249

RESULT 11

AAU91151

ID AAU91151 standard; Protein; 265 AA.

XX AC

AAU91151;

XX AC

DT 05-JUN-2002 (first entry)
 XX Streptomyces coelicolor pantothenate kinase Coax.
 DE
 XX
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.
 KW Streptomyces coelicolor.
 XX
 OS
 XX WO200216601-A2.
 PN
 XX 28-FEB-2002.
 PD
 XX 24-AUG-2001; 2001WO-US26531.
 PF
 XX 24-AUG-2000; 2000US-227860P.
 PR
 XX 20-MAR-2001; 2001US-0813453.
 PR
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA
 XX Yocum RR, Patterson TA;
 PI WPI; 2002-269358/31.
 XX
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 PS Claim 10; Page 69-70; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 265 AA;
 Query Match 49.2%; Score 629; DB 23; Length 265;
 Best Local Similarity 51.1%; Pred. No. 1.1e-57;
 Matches 134; Conservative 42; Mismatches 76; Indels 10; Gaps 5;
 1 MLLVIDGNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELF---RLAG--LGLD 55
 1 MLLTIDVGNTHVLGLFDGEDIVHWRISTDSRRTADELAVLLQGLMGHPLLLGDELG-D 59
 56 QIRAVLISSVVPPLTGLVLERLSLGVFGMRPLV-VGPGIKTGMPIQYDNPREGADRIYNA 114
 60 GIDGIAICATPVSVLHLEURTRRYGVDVPAVLVPGVKTGVPILTDHPKEVGADRIINA 119
 115 VAGKEYRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPLVISTEALFORASKLPRVDII 174
 120 VAARELYGGPAIVDFGTATTFDVAARGEYIGGVIAPIEISVALGVKAQLKIEVA 179
 175 RPSALIARTNSMQAGIYGVGLVDEIVTRMKAESKAP---RVIIATGGLASLIAPES 231
 180 RPRSVIGKNTVEAMQSGIYGVAGQVDCGVNRMARLADDDPDDVTVIATGGLAPVLGES 239
 232 KTIEAVEEYLLEGLRIILYERN 253
 240 SVIDHEPWLTLMLGLRIILYERN 261
 RESULT 12
 AAU91173
 ID AAU91173 standard; Protein; 260 AA.
 XX
 AC AAU91173;

XX 05-JUN-2002 (first entry)
 DT
 XX
 DE Pantcthenate kinase (Coax) #11.
 XX
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX
 OS Caulobacter crescentus.
 XX
 PN WO200216601-A2.
 PD
 XX 28-FEB-2002.
 PD
 XX 24-AUG-2001; 2001WO-US26531.
 PF
 XX 24-AUG-2000; 2000US-227860P.
 PR
 XX 20-MAR-2001; 2001US-0813453.
 PR
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA
 XX Yocum RR, Patterson TA;
 PI WPI; 2002-269358/31.
 XX
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 PS Claim 10; Page 102-103; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 260 AA;
 Query Match 43.9%; Score 561; DB 23; Length 260;
 Best Local Similarity 44.7%; Pred. No. 1.5e-50;
 Matches 114; Conservative 50; Mismatches 89; Indels 2; Gaps 1;
 1 MLLVIDGNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLQIRAV 60
 1 MLLAIEQGNNTNMFALHDGASWAQWRSATSTRDADEVVWLSQGLGFRADAV 60
 61 IISVVPPPLTGLVLERLSLGVFGMRPLVVGPGIKTGMPIQYDNPREGADRIYNAVAGYEK 120
 61 IISVVPOSIFENLRLNLSRYFNVEPLVIGENAKLGDIVRIEKPSEAGADRLVNAIGAAVY 120
 121 YTSLLIIVDFGTATTFDYVNRKGEYCGGAIAPLVISTEALFORASKLPRVDIIRSA-- 178
 121 YPGPLVWIDSGTATTFDVIADGAFEGGIIAEPINLSMQALHEAAKLPRIAIQRPAGNR 180
 179 IIRATVNSMQAGIYGVGLVDEIVTRMKAESKAPRVIIATGGLASLIAPESKTEAVE 238
 131 IVGTTVSAHQSGVFWGLISLIEGLVARIKAEERGEPTVIATGGVVASLEFEGATSDIDHFD 240
 239 EYLTLEGLRIILYERN 253
 241 SDLTIRGLLEIYRRN 255
 RESULT 13
 AAU91176
 ID AAU91176 standard; Protein; 219 AA.

AC AAU91152;
XX
XX DT 05-JUN-2002 (first entry)
XX
XX DE Mycobacterium tuberculosis pantothenate kinase Coax.
XX
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PN WO200216601-A2.
XX
XX PD 28-FEB-2002.
XX
XX PF 24-AUG-2001; 2001WO-US26531.
XX
XX PR 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX PI Yocum RR, Patterson TA;
XX
XX DR WPI; 2002-269358/31.
XX
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
XX PS Claim 10; Page 70-71; 128pp; English.
XX
XX CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX SQ Sequence 272 AA;
Query Match 39.8%; Score 509.5; DB 23; Length 272;
Best Local Similarity 39.6%; Pred. No. 4.1e-45;
Matches 105; Conservative 65; Mismatches 82; Indels 13; Gaps 5;
QY 1 MLLVVDGNTNIVLGIYDG----ERLVRDVRVSTDKARTDEYGLINELFRLAGLGLDQ 56
DB 1 MLLAIDVRNTHVTVGLLSGMKREHAKVQWQWRTSEVTADLALTIDGLI---GEDSER 57
QY 57 IRAVIISVVPLTGLVRLSLGFGMRP-LVVGPGIKTGMPIQYDNPREGVADRIVNAV 115
DB 58 LTGTAALSTVPSVLHEVRIMLDQYWPSPVPHVLIPEGVRTGIPLLVDNPKVCAIRVNC 117
QY 116 AGYEKRYSLIIVDGTATTFDYNRKGEYCGGATAGLVISTEALFQFASKLPRVDIIR 175
DB 118 AAYDRFRKAAIYDVGSSICVDVSAKGFEFLGGAIPGVQVSSDAAAARSAALRRVELAR 177
QY 176 PSAAIARTNVMQAGIYGYVGLVDEIVTRKAE-----SKDAP-RVIATGGGLASIAPE 230
DB 178 PRSVVGKNTVECMQAGVGFAGLVGLGRIDREYSGFSVDHDAIVATGHTAPILLPE 237
QY 231 SKTIEAVEEYLFLEGLRILYERNRE 255
DB 238 LHTVDHYDQHLTLQGLRLVFERNLE 262

Search completed: June 24, 2003, 21:46:15
Job time : 27.1803 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 22.343 Seconds
(without alignments)
2342.388 Million cell updates/sec

Title: US-09-813-453A-47
Perfect score: 1293
Sequence: 1 MILVIDGNTNTVLGVQDE.....DVIDSFLTKGLQLIKKNV 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertebrate.*
- 14: sp.unclassified.*
- 15: sp.rvirus.*
- 16: sp.bacteriap.*
- 17: sp.archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	1293	100.0	254	16	Q9KGH5	Q9KGH5 bacillus ha
2	990	76.6	258	2	Q9F985	Q9F985 bacillus st
3	849	65.7	259	16	Q8YAC5	Q8YAC5 listeria mo
4	848	65.6	259	16	Q92F54	Q92F54 listeria in
5	786	60.8	255	16	Q8R7M2	Q8R7M2 thermoanaer
6	785	60.7	259	16	Q8XHL5	Q8XHL5 clostridium
7	772	59.7	273	16	Q97EB4	Q97EB4 clostridium
8	662	51.2	265	16	Q9X8N6	Q9X8N6 streptomyce
9	598	46.2	261	16	Q9A6Z1	Q9A6Z1 caulobacter
10	494.5	38.2	274	16	Q9CD56	Q9CD56 mycobacteri
11	487.5	37.7	272	16	Q06282	Q06282 mycobacteri
12	451	34.9	256	16	Q8RFE4	Q8RFE4 fusobacteri
13	437.5	33.8	262	16	Q9RX54	Q9RX54 deinococcus
14	389.5	30.1	246	16	Q9WZ55	Q9WZ55 thermotoga
15	330.5	25.6	212	2	O32514	O32514 desulfovibr
16	320.5	24.8	273	16	O83446	O83446 treponema p

17	275.5	21.3	262	16	O51477	O51477 borrelia bu
18	193	14.9	56	2	P94305	P94305 bacillus ps
19	168	13.0	295	16	Q8Y2M4	Q8Y2M4 ralstonia s
20	155	12.0	276	16	Q8YQD7	Q8YQD7 anabaena sp
21	138.5	10.7	229	16	Q67753	Q67753 aquifex aeo
22	137.5	10.6	257	16	P74045	P74045 synechocyst
23	137	10.6	592	16	Q9JW17	Q9JW17 neisseria m
24	136	10.5	592	16	Q9JXF1	Q9JXF1 neisseria m
25	121	9.4	248	16	Q9HWC1	Q9HWC1 pseudomonas
26	110	8.5	209	16	Q9PIA9	Q9PIA9 campylobact
27	110	8.5	597	17	O30225	O30225 archaeoglob
28	107.5	8.3	223	16	Q9ZKY6	Q9ZKY6 helicobacte
29	107.5	8.3	242	16	Q9PC14	Q9PC14 xylella fas
30	99	7.7	224	16	Q98Q93	Q98Q93 mycoplasma
31	96.5	7.5	735	10	O9ZS89	Q9ZS89 arabidopsis
32	96.5	7.5	756	10	Q9M023	Q9M023 arabidopsis
33	95	7.3	1878	12	Q66580	Q66580 eastern equ
34	95	7.3	2493	12	Q88789	Q88789 eastern equ
35	94.5	7.3	313	16	Q97FW3	Q97FW3 clostridium
36	94.5	7.3	346	16	Q929W6	Q929W6 listeria in
37	94.5	7.3	855	17	Q973R5	Q973R5 sulfolobus
38	93.5	7.2	515	2	Q8RL85	Q8RL85 bacillus st
39	92.5	7.2	320	16	Q9JW56	Q9JW56 neisseria m
40	92.5	7.2	512	17	Q8TJY1	Q8TJY1 methanosarc
41	92	7.1	816	16	Q8R000	Q8R000 thermoanaer
42	91.5	7.1	512	16	Q8UBH3	Q8UBH3 agrobacteri
43	91	7.0	223	16	O25533	O25533 helicobacte
44	91	7.0	318	16	Q8ZAA3	Q8ZAA3 yersinia pe
45	91	7.0	435	16	Q9CH92	Q9CH92 lactococcus

ALIGNMENTS

RESULT 1

Q9KGH5 PRELIMINARY; PRT; 254 AA.
ID Q9KGH5;
AC Q9KGH5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
Horikoshi K.;
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AF001507; BAB03805.1; -;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; OE3B557BA7DAC176 CRC64;

Query Match 100.0%; Score 1293; DB 16; Length 254;
Best Local Similarity 100.0%; Pred. No. 5e-108;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MILVIDGNTNTVLGVQDETLVHHWRLATSROKTEDEYAMTVRSFLFDHAGLFQDIDGI 60
|||||
DB 1 MILVIDGNTNTVLGVQDETLVHHWRLATSROKTEDEYAMTVRSFLFDHAGLFQDIDGI 60
QY 61 VISSVPPPMFSLQMKCKYFHVTPMIIIGPGIKTGLNIKNYDNPKEVGADRIVNAVAIAEL 120

Db 61 VISSVPPPMFSLQCKKFFHYTPMIGPKTGLNKNYDNPKEGVADRVNAVAIEL 120
QY 121 YGYPAIVDFGTATTYCLINEKQYAGGVYAPGIMISTEALYHRASKLPRIETAKPKQV 180
Db 121 YGYPAIVDFGTATTYCLINEKQYAGGVYAPGIMISTEALYHRASKLPRIETAKPKQV 180
QY 181 GTNTIDSMQSGIFYGVYQVGVVKKMAQAESEPKVIATGGTGLAKLIGTSETIDVDSF 240
Db 181 GTNTIDSMQSGIFYGVYQVGVVKKMAQAESEPKVIATGGTGLAKLIGTSETIDVDSF 240
QY 241 LTLKGLQLIYKKNV 254
Db 241 LTLKGLQLIYKKNV 254

RESULT 2

Q9F985 PRELIMINARY; PRT; 258 AA.
ID Q9F985
AC Q9F985
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative 32 kDa replication protein.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF198621; AAC28531.1;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 76.6%; Score 990; DB 2; Length 258;
Best Local Similarity 71.3%; Pred. No. 8.2e-81;
Matches 181; Conservative 35; Mismatches 38; Indels 0; Gaps 0;
QY 1 MILVIDGNTNTVLGYQDETLLVHWRATSRKTEDEYAMTVRSFLDHAGLQFQIDGI 60
Db 1 MIFVLDGNTNTVLGYQDGLKHKWRIETSRKTEDEYGMTIKALLNHVGLQFSDIGI 60
QY 61 VISSVPPPMFSLQCKKFFHYTPMIGPKTGLNKNYDNPKEGVADRVNAVAIEL 120
Db 61 IISVVPPPMFSLQCKKFFHYTPMIGPKTGLNKNYDNPKEGVADRVNAVAIEL 120
QY 121 YGYPAIVDFGTATTYCLINEKQYAGGVYAPGIMISTEALYHRASKLPRIETAKPKQV 180
Db 121 YGSPLIVDFGTATTYCLINEKQYAGGVYAPGIMISTEALYHRASKLPRIETAKPKQV 180

QY 181 GTNTIDSMQSGIFYGVYQVGVVKKMAQAESEPKVIATGGTGLAKLIGTSETIDVDSF 240
Db 181 GNTVSAMQAGILYGVYQVGVVKKMAQAESEPKVIATGGTGLAKLIGTSETIDVDSF 240
QY 241 LTLKGLQLIYKKNV 254
Db 241 LTLKGLQLIYKKNV 254

RESULT 3

Q8YAC5 PRELIMINARY; PRT; 259 AA.
ID Q8YAC5
AC Q8YAC5
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.

OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
Raquero F., Berche P., Bloeker H., de Daruvar A., Dehoux P.,
Charbit A., Chetoui F., Couve E., de Daruvar A., Durand L.,
Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL591974; CAD00748.1;
DR MEROPS; M41.009;
DR ListList; LMO00221;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 65.7%; Score 849; DB 16; Length 259;
Best Local Similarity 60.1%; Pred. No. 3.8e-68;
Matches 152; Conservative 50; Mismatches 51; Indels 0; Gaps 0;
QY 1 MILVIDGNTNTVLGYQDETLLVHWRATSRKTEDEYAMTVRSFLDHAGLQFQIDGI 60
Db 1 MILVIDGNTNTVLGYQDETLLVHWRATSRKTEDEYAMTVRSFLDHAGLQFQIDGI 60
QY 61 VISSVPPPMFSLQCKKFFHYTPMIGPKTGLNKNYDNPKEGVADRVNAVAIEL 120
Db 61 IISVVPPPMFSLQCKKFFHYTPMIGPKTGLNKNYDNPKEGVADRVNAVAIEL 120
QY 121 YGYPAIVDFGTATTYCLINEKQYAGGVYAPGIMISTEALYHRASKLPRIETAKPKQV 180
Db 121 YGTVPIVDFGTATTYCLINEKQYAGGVYAPGIMISTEALYHRASKLPRIETAKPKQV 180
QY 181 GTNTIDSMQSGIFYGVYQVGVVKKMAQAESEPKVIATGGTGLAKLIGTSETIDVDSF 240
Db 181 GKSTVSSMQAGIFYGVYQVGVVKKMAQAESEPKVIATGGTGLAKLIGTSETIDVDSF 240
QY 241 LTLKGLQLIYKKNV 253
Db 241 LTLKGLQLIYKKNV 253

RESULT 4

Q92F54 PRELIMINARY; PRT; 259 AA.
ID Q92F54
AC Q92F54
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lin0253.
GN LIN0253.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

```

Best Local Similarity   56.5%; Pred. No. 1.7e-62;
Matches 143; Conservative    49; Mismatches 61; Indels      0; Gaps      0;

QY  1 MILVIDGNTNTVLGVVQDFTLVHHWRLATSRKOTDEYAMTVRSLPDHAGLQFDDIDGI 60
Db  1 MLIAFDVGNTNIWVGFKGKKLLHSFRISTDKNKTYDEYGLMVLNQLIGNGISLTIEDDV 60
QY  61 VISSVPPMFSLSEOMCKKYFHVTPTMIIGPGIKTGLNIKVDNPKEVGADRIVNAVAIEL 120
Db  61 IISVVPLMTNLQVMNSLKVFRTKPVIIVGPFGITGINIKVDNPKEVGADRIVNAVAAYEL 120
QY  121 YGPAIWDGTATTTCYLINKEKYOYAGGVTAAPGMISTEALYHRASKLPRIETAKPKQVV 180
Db  121 YGGPVIWDGTATTFCAISEKGLEYLGIIAPGLMISADALFORTAKLPRLIDLTKPPTVI 180
QY  181 GTNTDSMQSGIFYGVSQDVGYYVKRMAQAASEPKVIATGGKLKLGTESETIDVTDTSF 240
Db  181 NRNVASMSQGIIGHVGMVDYIVTRMKGEFAFSYVVATGGFANMLAESKTIDTVNEM 240
QY  241 LTLKGLQLIYKKN 253
Db  241 LTLEGLRIIYERN 253

RESULT 6
Q8XHL5 PRELIMINARY; PRT; 259 AA.
AC Q8XHL5;
DT 01-MAR-2002 (TrEMBRel. 20, Created)
DI 01-MAR-2002 (TrEMBRel. 20, Last sequence update)
DO 01-JUN-2002 (TrEMBRel. 21, Last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB82174.1; -.
DR InterPro; IPRO04619; baf.
DR InterPro; IPRO00515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
SQ Hypothetical protein; Complete proteome.
KW SEQUENCE 259 AA; 28819 MW; OD5FEA3B/AJ45EI0 CRC64;

Query Match          60.7% ; Score 785; DB 16; Length 259;
Best Local Similarity 58.0% ; pred. No. 2.1e-62;
Matches 149; Conservative    48; Mismatches 56; Indels      4; Gaps      1;

QY  1 MILVIDGNTNTVLGVVQDFTLVHHWRLATSRKOTDEYAMTVRSLPDHAGLQFDDIDGI 60
Db  1 MILLIDVGNTNVLGIHNDNEKIYASWRISDSKTSDEYSIQVMOLFNNQALPNEDVEGI 60
QY  61 VISSVPPMFSLSEOMCKKYFHVTPTMIIGPGIKTGLNIKVDNPKEVCADRIVNAVAIEL 120
Db  61 IISSVPNIMHLSLENMVVRKCFCKEPIVVGPGITGINIKVDNPKEVGADRIVNAVAFEK 120
QY  121 YGPAIWDGTATTTCYLINKEKYOYAGGVTAAPGMISTEALYHRASKLPRIETAKPKQVV 180
Db  121 HKPFMIIDGTATTFCAITEKG DYLGNCICPGIOISADALFERAAKLPRIELEKPSVI 180
QY  131 GTNTIDSQSQGFYGVSQDVGYYVKRMAQ-----AESEPKVATGGKLKLGITESETIDV 236

```

```
Db 181 CKNTVTSMQAGIIGYIGKVEYIVKRMKEMMDLGEKEPFVLATGGLAKLVYSETVIDE 240
QY 237 IDSFLTLKGLQLIYKKN 253
Db 241 VDRKLTLEGLKILYKKN 257

RESULT 7
Q97EB4 ID Q97EB4 PRELIMINARY; PRT; 273 AA.
AC Q97EB4; 2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory
DE factor.
GN CAC3200.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
NCBI_TaxID=1488;
X [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AFCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007815; AAK81136.1; -
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD.transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANS_PNN_MEMBER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;

Query Match 59.7%; Score 772; DB 16; Length 273;
Best Local Similarity 56.4%; Pred. No. 3.3e-61;
Matches 145; Conservative 47; Mismatches 61; Indels 4; Gaps 1;

QY 1 MILVIDGNTNVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
Db 12 VILVLDVGNNTVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
Db 61 VISSVPPMMFSLQMKCKYFHYVTPMIIGPGIKTGLNIRYDNPKEVGADRIYNAVAIEL 120
Db 72 IISVVPPNIMSLHMRKYKINPLVVGPGIKTGINIRYDNPKEVGADRIYNAVAHEI 131
QY 121 YGPAIVDFGATATTCCLINEKQYAGGVIAPGIMISTALYHRASKLPRIETAKPKQV 180
Db 132 YKRLIIDFGATATTCFAVRGNDYLGGAICPGIKVSSEALFEKAALPRVELIKPAYAI 191
QY 181 GTNTIDSMQSGIFGVVSDGVVVKRMK---AQASEPKVATGGLAKLIGTSTIDY 236
Db 192 CKNTISQSGIVGVYQGVYIVRMKEELOEGEKEPLVWATGGLAKLIGTSTIDY 251
QY 237 IDSFLTLKGLQLIYKKN 253
Db 252 INFFLEGLRIIYKKN 268

RESULT 8
Q9X8N6 ID Q9X8N6 PRELIMINARY; PRT; 265 AA.
AC Q9X8N6; 1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
```

```
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapajite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL049628; CAB40880.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 51.2%; Score 662; DB 16; Length 265;
Best Local Similarity 49.8%; Pred. No. 2.4e-51;
Matches 132; Conservative 55; Mismatches 64; Indels 14; Gaps 4;

QY 1 MILVIDGNTNVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQ 53
Db 1 MLITIDVGNTHVLGLFDGEDIVERHRISTRITSDRTADELAVLQGLMGHPLIGDELG-- 58
QY 54 FODIDIGIVSSVVPMMFSLQMKCKYFHYVTP--MIIGPGIKTGLNIRYDNPKEVGADRI 112
Db 59 -DGIDGIALCAIVPSVLHRELVRTRYRGVDFAVLVEPGVTKGVPILTDHPKEVGADRI 117
QY 113 NAAVAELGYPAIVVDFGATATTCCLINEKQYAGGVIAPGIMISTALYHRASKLPRIE 172
Db 118 NAAVAELGYPAIVVDFGATATTCCLINEKQYAGGVIAPGIMISTALYHRASKLPRIE 177
QY 173 IAKPKQVQVGTNTIDSMQSGIFGVVSDGVVVKRMKAQASEP---KVIATGGLAKLIGT 229
Db 178 VARPRSVIGKNTVEAMQSGIVGFAGQVDGVVNRMARLADDDPDVTVIATGGLAPMWLG 237
QY 230 ESETIDVIDSFLTLKGLQLIYKKN 254
Db 238 ESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 9
```

Q9A6Z1	Q9A6Z1	PRELIMINARY;	PRT;	261 AA.
AC	Q9A6Z1	Q9A6Z1		
DT	01-JUN-2001	(TrEMBLrel. 17, Created)		
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)		
DE	Transcriptional activator, putative, Baf family.			
GN	CC1935.			
OS	Caulobacter crescentus.			
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;			
OC	Caulobacter.			
OX	NCBI_TaxID=155892;			
EN	SEQUENCE FROM N.A.			
RP	STRAIN=ATCC 19089 / CB15;			
RC	MEDLINE=21173698; PubMed=11259647;			
RX	Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,			
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,			
RA	Potocka I., Nelson W.C., Durkin A.S., Stephens C., Phadke N.D., Ely B.,			
RA	Deboy R.T., Dodson R.J., Newton A.S., Gwinn M.L., Haft D.H.,			
RA	Kolony J.F., Smith J., Craven M.B., Khouri H., Shetty J., Berry K.,			
RA	Utterback T., Tran K., Wolf A., Vamathevan J., Hermodson J.,			
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,			
RT	"Complete genome sequence of Caulobacter crescentus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).			
DR	EMBL; AE005867; AA23910.1; -			
DR	TIGR; CC1935; -			
DR	InterPro; IPR004619; Baf.			
DR	Pfam; PF03309; Bvg_acc.factor; 1.			
DR	TIGRFAMS; TIGR00671; baf; 1.			
KW	Complete proteome.			
QY	SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;			
Query Match	46.2%; Score 598; DB 16; Length 261;			
Best Local Similarity	43.9%; Pred. No. 1.3e-45;			
Matches	112; Conservative 61; Mismatches 80; Indels 2; Gaps 1;			
QY	1 MILVIDGNTNTVLGVYQDETLLVHHWRLATSRKTEDEYAMTVRSFLDHAGLQFDIGI 60			
Db	2 MLLAIEGQNTNTMFAIHGASWVAQWRSATSTRTADEYVWLLSQLLSMQGLGFRADAV 61			
QY	61 VISSVPPMFLSEOMCKKFFHYTPMIIIGIKTGLNKKYDNKPKVGGADRVNVAIAEL 120			
Db	62 IISVVPPQISINLNRNRRVFNVEPLVIGNAKIGIDVRIEKPSEAGADRLVNAIGAAMV 121			
QY	121 YGPAIVVDFTATTYCLINEKKQYAGGVIAPIAGIMISTEALYHRASKLPRIETAKP--KQ 178			
Db	122 YGPLVVIDSGTATTFDIAADGAFEGGIATPGLNLSQALHEAAAKLPRIATQRPAGNR 181			
QY	179 VVGNTTIDSMQSGIFYGVSDVGVVKKMAQAESEPKVIATGSLAKLIGETSIDVID 238			
Db	182 IGVGDTVSAMQSGVFWYISLIEGLVARKEAERPEMTVIATGVSALFEGATDSIDHFD 241			
QY	239 SFLTLGLQLIYKKN 253			
Db	242 SLLTIRGLLEIYRKN 256			
RESULT 10				
Q9CD56	Q9CD56	PRELIMINARY;	PRT;	274 AA.
AC	Q9CD56	Q9CD56		
DT	01-JUN-2001	(TrEMBLrel. 17, Created)		
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)		
DE	Hypothetical protein ML0232.			
GN	ML0232.			
OS	Mycobacterium leprae.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1769;			
RP	SEQUENCE FROM N.A.			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

DM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 10.3055 Seconds
(without alignments)
725.188 Million cell updates/sec

Title: US-09-813-453A-47

Perfect score: 1293

Sequence: 1 MILVIDGNTNTVLGVYQDE.....DVIDSFLTLKGLIYKKNV 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*

6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81.5	6.3	1529	4	US-09-134-001C-3945
2	80	6.2	343	2	US-08-599-171A-28
3	80	6.2	343	2	US-08-646-590B-28
4	80	6.2	343	3	US-09-069-226-28
5	80	6.2	343	4	US-09-412-184-28
6	79.5	6.1	1350	2	US-08-319-866-9
7	78	6.0	255	4	US-08-815-225-4
8	77	6.0	330	4	US-09-134-001C-3582
9	77	6.0	566	3	US-09-264-737-2
10	76	5.9	333	4	US-09-134-001C-3946
11	76	5.9	4550	2	US-08-804-227C-8
12	76	5.9	4550	2	US-08-804-198-2
13	75.5	5.8	679	1	US-08-441-139-5
14	74.5	5.8	877	1	US-08-430-925A-4
15	73.5	5.7	639	1	US-08-466-390-2
16	73.5	5.7	639	1	US-08-470-950-2
17	73.5	5.7	639	1	US-08-467-781-2
18	73.5	5.7	639	1	US-08-195-487-2
19	73.5	5.7	639	2	US-08-483-924-2
20	73.5	5.7	639	5	PCT-US93-06160-2
21	73	5.6	539	3	US-09-264-737-1
22	73	5.6	573	2	US-08-706-209-1
23	73	5.6	573	3	US-08-981-787-1
24	73	5.6	573	4	US-08-461-722-1
25	73	5.6	573	4	US-08-336-251-1
26	73	5.6	573	5	PCT-US94-06362-1
27	73	5.6	573	5	PCT-US96-11373-1

28 73 5.6 573 5 PCT-US96-11375-1 Sequence 1, Appl
29 72.5 5.6 301 2 US-08-716-301-2 Sequence 2, Appl
30 72.5 5.6 775 1 US-07-966-278-1 Sequence 1, Appl
31 72.5 5.6 775 1 US-08-424-921-1 Sequence 1, Appl
32 72.5 5.6 775 2 US-08-556-355A-1 Sequence 1, Appl
33 72.5 5.6 775 2 US-07-803-627A-1 Sequence 37, Appl
34 72.5 5.6 776 2 US-08-688-649-37 Sequence 5339, Ap
35 72 5.6 212 4 US-09-134-001C-5339 Sequence 8, Appl
36 72 5.6 333 4 US-09-562-737-8 Sequence 2, Appl
37 72 5.6 372 2 US-08-846-762-79 Sequence 79, Appl
38 72 5.6 562 4 US-09-227-717-2 Sequence 2, Appl
39 72 5.6 562 4 US-09-173-151A-22 Sequence 22, Appl
40 72 5.6 613 4 US-09-446-504-5 Sequence 5, Appl
41 72 5.6 613 4 US-09-712-266-5 Sequence 5, Appl
42 72 5.6 613 4 US-09-091-889A-2 Sequence 2, Appl
43 72 5.6 1037 4 US-09-134-001C-4794 Sequence 4794, Ap
44 71.5 5.5 285 4 US-09-327-681-6 Sequence 6, Appl
45 71.5 5.5 740 4 US-09-323-872A-25 Sequence 25, Appl

ALIGNMENTS

RESULT 1

US-09-134-001C-3945

; Sequence 3945, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3945

; LENGTH: 1529

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3945

Query Match 6.3%; Score 81.5; DB 4; Length 1529;

Best Local Similarity 21.4%; Pred. No. 5.4;

Matches 50; Conservative 30; Mismatches 73; Indels 81; Gaps 9;

QY 2 ILVIDVGNNTVLGVYQDETLVHHWRLATSRQKTEDEYMTVRSFLDHAGLQFQDIDGIV 61

Db 635 ILVLD-----DTSLTH-----ENSTAM----- 651

QY 62 ISSVPPPMFSLQCKKRYFHVTPMIGIKTNLKNPKYKGVADRVNVAIAELY 121

Db 652 -----PILLALS-----HVHQL-----IREGLMETSLTIAQSGETREHVHVACLLGY 694

QY 122 GYPAIVDFGATYTYCLINEKKQYAGGVIAIGIMISTEALYHRASKLPRIETAKPKQVVG 181

Db 695 GANAVPYLAORTIEQLTROGO--LSGTVAENVATYTNVLSGVTKV-----MAK-----MG 744

QY 182 TITIDSMQSGIFYGVYQDVGVVKKMAQAESEPKVIATGGLAKLIGTSETID 235

Db 745 ISTVQSYGAQIFPAVGLNSVIEK-----YFTGTQSKLSGISIEQID 787

RESULT 2

US-08-599-171A-28

; Sequence 28, Application US/08599171A

; Patent No. 5814473

; GENERAL INFORMATION:

; APPLICANT: WARREN, Patrick V.

; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES

NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,171A
FILING DATE: Concurrently
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HERRON, CHARLES J.
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-38
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-599-171A-28

Query Match 6.2%; Score 80; DB 2; Length 343;
Best Local Similarity 22.4%; Pred. No. 0.75;
Matches 59; Conservative 31; Mismatches 53; Indels 120; Gaps 14;
QY 46 LFDHAGLQD-IDGIVISSV---VPPMFSLEQCKKYFHYTPMIGIKTGLNIKYD 101
DB 23 LYDEEGKYLDFVSGIGVNSLGHAYPKLTKALKEQVEKLLHVSNL-----YE 69
QY 102 NP-KEVGADRV-----NAVAATL-----120
DB 70 NPQBELAHLKLVHFWTEGKVFANSCTESVEAAIKLARKYWRDKGNKWKFKISFNSFH 129
121 -----YGYPAIVDFGTA-----TTYCLINEKK-----QYAGGYI 150
130 GRTYGSLSATGQPKFKHGFPLVPGFSYAKLNDIDSVKLLDEETAGIIIEVIQGGV- 188
QY 151 APGIMISTALYHRASKLPRIETAKPKQVGVNTIDSMOSGI-----FGYYSQVDGVYK 205
DB 189 -----NEASEDFLSKLQ--EICKEDVL--LIIDEVQTGIGTGEFYAY-----228
QY 206 RMKAQAESEPKVIATGGLAKLIG 228
DB 229 ---QHFNLPDVI---LAKGLG 245

RESULT 3
US-08-646-590B-28
Sequence 28, Application US/08646590B
Patent No. 5962283
GENERAL INFORMATION:
APPLICANT: Warren, Patrick V.
APPLICANT: Swanson, Ronald V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P. C.

STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,590B
FILING DATE: 08-May-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,171
FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01094
FILING DATE: 21-January-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-646-590B-28

Query Match 6.2%; Score 80; DB 2; Length 343;
Best Local Similarity 22.4%; Pred. No. 0.75;
Matches 59; Conservative 31; Mismatches 53; Indels 120; Gaps 14;
QY 46 LFDHAGLQD-IDGIVISSV---VPPMFSLEQCKKYFHYTPMIGIKTGLNIKYD 101
DB 23 LYDEEGKYLDFVSGIGVNSLGHAYPKLTKALKEQVEKLLHVSNL-----YE 69
QY 102 NP-KEVGADRV-----NAVAATL-----120
DB 70 NPQBELAHLKLVHFWTEGKVFANSCTESVEAAIKLARKYWRDKGNKWKFKISFNSFH 129
121 -----YGYPAIVDFGTA-----TTYCLINEKK-----QYAGGYI 150
130 GRTYGSLSATGQPKFKHGFPLVPGFSYAKLNDIDSVKLLDEETAGIIIEVIQGGV- 188
QY 151 APGIMISTALYHRASKLPRIETAKPKQVGVNTIDSMOSGI-----FGYYSQVDGVYK 205
DB 189 -----NEASEDFLSKLQ--EICKEDVL--LIIDEVQTGIGTGEFYAY-----228
QY 206 RMKAQAESEPKVIATGGLAKLIG 228
DB 229 ---QHFNLPDVI---LAKGLG 245

RESULT 4
US-09-069-226-28
Sequence 28, Application US/09069226
Patent No. 6013509
GENERAL INFORMATION:
APPLICANT: WARREN, Patrick V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN.

```

; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/069,226
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,171
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HERRON, CHARLES J.
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-38
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 AMINO ACIDS
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-069-226-28

```

```

Query Match      6.2%; Score 80; DB 3; Length 343;
Best Local Similarity 22.4%; Pred. No. 0.75;
Matches 59; Conservative 31; Mismatches 53; Indels 120; Gaps 14;

Qy 46 LFDHAGLQFQD-IDGIVISSV---VPPMMFSLQMKCKYFHTVPMIIGPGIKTGLNICYD 101
Db 23 LYDEGREYLDVFSGIGVNSLGHAYPKLTEALKEQVEKLLHVSNL-----YE 69
Qy 102 NP-KEVGADRV-----NAVAATEL----- 120
Db 70 NPQOEELAHKLKVFHWTEGKVFVANSGETESVEAAIKLARKYWRDKGNKWKFI SFNSFH 129
Qy 121 -----YGYPAIVVDFGTA-----TTTYCLINEKK-----QYAGGVI 150
Db 130 GRTYGSLSATQPKFHKGFEPVPGFSYAKLNDIDSVYKLLDEETAGIIIEVIOGEGV- 188
Qy 151 APGIMISTEALYHRASKLPRIETIAKPKOVVGTNTIDSMQSGI-----FYGVVSQVDGVVK 205
Db 189 -----NEASEDFLSKIQ--EICKEDVL--LIIDEVQTGIGRTGEFYAY----- 228
Qy 206 RMKAQAESEPKVIATGGLAKLIG 228
Db 229 ---QHFNKLPDVIA---LAKGLG 245

```

```

RESULT 5
US-09-412-184-28
; Sequence 28, Application US/09412184
; Patent No. 6268188
; GENERAL INFORMATION:
; APPLICANT: Warren, Patrick V.
; APPLICANT: Swanson, Ronald V.
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/412,184
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,590
; FILING DATE: 08-May-1996
; APPLICATION NUMBER: 08/599,171
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/01094
; FILING DATE: 21-January-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/017001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-412-184-28

```

```

Query Match      6.2%; Score 80; DB 4; Length 343;
Best Local Similarity 22.4%; Pred. No. 0.75;
Matches 59; Conservative 31; Mismatches 53; Indels 120; Gaps 14;

Qy 46 LFDHAGLQFQD-IDGIVISSV---VPPMMFSLQMKCKYFHTVPMIIGPGIKTGLNICYD 101
Db 23 LYDEGREYLDVFSGIGVNSLGHAYPKLTEALKEQVEKLLHVSNL-----YE 69
Qy 102 NP-KEVGADRV-----NAVAATEL----- 120
Db 70 NPQOEELAHKLKVFHWTEGKVFVANSGETESVEAAIKLARKYWRDKGNKWKFI SFNSFH 129
Qy 121 -----YGYPAIVVDFGTA-----TTTYCLINEKK-----QYAGGVI 150
Db 130 GRTYGSLSATQPKFHKGFEPVPGFSYAKLNDIDSVYKLLDEETAGIIIEVIOGEGV- 188
Qy 151 APGIMISTEALYHRASKLPRIETIAKPKOVVGTNTIDSMQSGI-----FYGVVSQVDGVVK 205
Db 189 -----NEASEDFLSKIQ--EICKEDVL--LIIDEVQTGIGRTGEFYAY----- 228
Qy 206 RMKAQAESEPKVIATGGLAKLIG 228
Db 229 ---QHFNKLPDVIA---LAKGLG 245

```

```

RESULT 6
US-08-319-866-9
; Sequence 9, Application US/08319866
; Patent No. 5929223
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy P.
; APPLICANT: Yin, Jerry C.
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
; TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive

```

CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,866
FILING DATE: 7-OCT-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL94-03
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1350 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-319-866-9

Query Match 6.1%; Score 79.5; DB 2; Length 1350;
Best Local Similarity 20.4%; Pred. No. 7.6;
Matches 57; Conservative 35; Mismatches 85; Indels 103; Gaps 15;

QY 4 VIDGNTVNLGVYQDETIVHWRLATSR-----QKTE--DEYAMTVRSFLDHA 50
Db 340 VFCRVVTTSGMF--EAICNHKIKATNKGRLRSALTIFPQRTDAKHRYRNWNLISYA 397
QY 51 GLQFQDIDGIVISSVVPMMFSLQMKCK-----FHVTP 85
Db 398 G--YKQADGKIIGD---PMNVFTEVCTKLWKSKEWDILPLVVSANGHDPDYDYP 452
QY 86 MIIGPGIKTKLNIKYDNPK-EVGADRVNAVAIAELGYPAI---VVDG-----TATYV- 136
Db 453 ELI-----LEVPLTHPRFEWSD-----LGLRWYALPAVSSMLFDVGGIQTATTS 499
QY 137 -----CLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQVVGTVN 183
Db 500 GWTMSPEIGSRNLCDTNRN-----MLETVALKMLDTRPTPSLWKKAVV--- 545
QY 184 TIDSMQSGIFYGVSVQDGVVKKRKAQA-----ESEPKV 217
Db 546 ----EMNIALVHSQSRNVTIVDHHHTASEFSMKHFENESKL 582

RESULT 7
US-08-815-225-4
Sequence 4, Application US/08815225
Patent No. 6268479
GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Yao, Shi Du
TITLE OF INVENTION: AN INTRACELLULAR AMYLOID-BETA PEPTIDE
TITLE OF INVENTION: BINDING (ERAB) POLYPEPTIDE AND METHODS FOR INHIBITING
TITLE OF INVENTION: NEURODEGENERATIVE CONDITIONS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: NY

CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,866
FILING DATE: 7-OCT-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL94-03
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1350 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-319-866-9

Query Match 6.1%; Score 79.5; DB 2; Length 1350;
Best Local Similarity 20.4%; Pred. No. 7.6;
Matches 57; Conservative 35; Mismatches 85; Indels 103; Gaps 15;

QY 4 VIDGNTVNLGVYQDETIVHWRLATSR-----QKTE--DEYAMTVRSFLDHA 50
Db 340 VFCRVVTTSGMF--EAICNHKIKATNKGRLRSALTIFPQRTDAKHRYRNWNLISYA 397
QY 51 GLQFQDIDGIVISSVVPMMFSLQMKCK-----FHVTP 85
Db 398 G--YKQADGKIIGD---PMNVFTEVCTKLWKSKEWDILPLVVSANGHDPDYDYP 452
QY 86 MIIGPGIKTKLNIKYDNPK-EVGADRVNAVAIAELGYPAI---VVDG-----TATYV- 136
Db 453 ELI-----LEVPLTHPRFEWSD-----LGLRWYALPAVSSMLFDVGGIQTATTS 499
QY 137 -----CLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQVVGTVN 183
Db 500 GWTMSPEIGSRNLCDTNRN-----MLETVALKMLDTRPTPSLWKKAVV--- 545
QY 184 TIDSMQSGIFYGVSVQDGVVKKRKAQA-----ESEPKV 217
Db 546 ----EMNIALVHSQSRNVTIVDHHHTASEFSMKHFENESKL 582

RESULT 7
US-08-815-225-4
Sequence 4, Application US/08815225
Patent No. 6268479
GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Yao, Shi Du
TITLE OF INVENTION: AN INTRACELLULAR AMYLOID-BETA PEPTIDE
TITLE OF INVENTION: BINDING (ERAB) POLYPEPTIDE AND METHODS FOR INHIBITING
TITLE OF INVENTION: NEURODEGENERATIVE CONDITIONS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: NY

COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/815,225
FILING DATE: 12-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION/DOCKET NUMBER: 0575/55209
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 255 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-815-225-4

Query Match 6.0%; Score 78; DB 4; Length 255;
Best Local Similarity 21.5%; Pred. No. 0.8;
Matches 50; Conservative 34; Mismatches 75; Indels 74; Gaps 10;

QY 36 EDEYAMTVRSFLDHAQLQ-----FODIDGIVISSVVPMMFSLQ 75
Db 40 DEEGATARELGDARQYQHLDTIEDWQVRVAYAREEFGSDGLVNNAGISTGMFLETE 99
QY 76 MCKGYFHYTPMIGPGIKTKLNIKYDNPKVAGADRVNAVAIAELGYPAIVVDFGTAT 135
Db 100 SVFERFKVVDIINL-TGVFIGMKTVIPAMKDAGGGSIVNISSAAGLMGL-ALTSSYGAS-- 155
QY 136 YCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQVVGTVNII--DSMQSGIF 193
Db 156 -----RWGVRGL-----SKLAAYEL-----GTDRIRVNSVHPGMT 185
QY 194 YGVSVQDGVVKKRKAQA-----SEPKVIATGGLAKLIGTSETI 234
Db 186 YTPMIAETGI-----RQSGNYPTPMGRVGNPEGEIA-GAVVKLLSDTSSVY 232

RESULT 8
US-09-134-001C-3582
Sequence 3582, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3582
LENGTH: 330
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3582

Query Match 6.0%; Score 77; DB 4; Length 330;
Best Local Similarity 16.5%; Pred. No. 1.6;
Matches 55; Conservative 39; Mismatches 95; Indels 144; Gaps 12;

QY 1 MILVIDGNTNVLGVY-QDEFLVHHWRLATSRQKTEDEYAMTVRSLEFDHAG-----51
Db 6 ILIAADIGCTCKLGIDFKDLQHLKWSIDTDT-----DHTGELLKNYI 51
QY 52 -----LQFODIDGIVISSVVP-----PMFSLQCKKYFHYT 84
Db 52 NSETEKIAEKYKDFNVGVGVGPGVDFDGVVYGVANLHWPSVNVREIFKQYVACP 111
QY 85 PMIIIGPKTGKLNKIDNPKEVGADRIVNAVAIELYGPALVDFGTATTYCLINEKKQ 144
Db 112 VYVDNDANVAALGEKHGAGE-GADDVV-----AITLGTG-----145
QY 145 YAGGVIAPIGMI-----STEALYHRAS-----166
Db 146 LGGGIISNGEIVHGHNGSGAEIGHLRADFQRCFCNCGKSGCIETVASATGVVNLVNFY 205
QY 167 -----KLPIRIETAKPKOVVGTNTIDMSQSG-----IFYGVYSQVDGVVKRKA 209
Db 206 PKLTFKSSILQIKDNQVAKAVFADAAKAGDQFCIFITEKVANYIGYLCSSIISVT-----260
QY 210 QAESEPKVIATGGLAKLIGTSETIDVIDSFLT 242
Db 261 ----SNPKYIVLGG-----GMSTAGLILLENIKT 285

RESULT 9

US-09-264-737-2
; Sequence 2, Application US/09264737A
; Best Local Similarity 20.3%; Pred. No. 3.7;
; Patent No. 6107549
; GENERAL INFORMATION:
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Ruff, Thomas G.
; TITLE OF INVENTION: Engineering Plant Resistance to Pyridines via
; FILE OF INVENTION: Expression of Esterase Enzymes
; FILE REFERENCE: 38-21(10551) RLE3 Pyridine Tolerance
; CURRENT APPLICATION NUMBER: US/09/264,737A
; CURRENT FILING DATE: 1999-03-09
; EARLIER APPLICATION NUMBER: 60/077,377
; EARLIER FILING DATE: 1998-03-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Rabbit
US-09-264-737-2

Query Match 6.0%; Score 77; DB 3; Length 566;
Best Local Similarity 20.3%; Pred. No. 3.7;
Matches 47; Conservative 34; Mismatches 65; Indels 86; Gaps 12;
QY 3 LVIDGNTNVLGVYQDEFLVHHWRLATSRQKTEDE-YAMTVRSLEFDHAG-----51
Db 268 TAEIAGCKTTTSV-----MWH-----CLRKTEELMEVLTKMKFMDLVGDPKENTA 317
QY 52 LQFODIDGIVISSVVPMMFSLQCKKYFHYTPIIIGPKTGKLNKIDNPKEVGADRI 111
Db 318 FLTTVIDGVLLPKA-----PAEILAEKKYNMLPYMG-----INQDEFG-----356
QY 112 VNVAIAELYGPALVDFGTAT-----TYCLINEKKQ-----YAGGVIAPI-----152
Db 357 ----WIIPQMGLYPLSEGKLDOKTATELLKWSYPIVNVSKELTPVATEKYLGGTDPVKKK 414
QY 153 -----GIMISTEALYHRASKLPRI-----EIAKPKOVVG 181
Db 415 DLFLDNLADLLFGVPSVNVARHHRDAGAPTMYEYRPSFSSDMRPKTVIG 466

RESULT 10

US-09-134-001C-3946
; Sequence 3946, Application US/09134001C
; Patent No. 6380370

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3946
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3946

Query Match 5.9%; Score 76; DB 4; Length 333;
Best Local Similarity 20.2%; Pred. No. 2.1;
Matches 49; Conservative 44; Mismatches 69; Indels 80; Gaps 14;
QY 25 HWRLATSRQKTEDE-YAMTVRSLEFDHAGLQFODIDGIVISSVVPMMF-----71
Db 53 HW--ADEODTSDLAYNASIKAIED-AGIQPVDIDMIIVATATGDMPPFVSANILQERLG 119
QY 72 -----SLEOM--CKKYFHVTPMIIGPG-IKTGLNLIKYNPKKEVGADRIVNAVAIAELYGY 123
Db 120 TGVKATMDQLAACSGFMV--SMITAKQYIQSG---DYKHLVVGADKL-----162
QY 124 PAIVVDFTGATTYCLINEKKQYAGVIAPIGMISTEALYHRASKLPRIEIAKPKOVVGTN 183
Db 153 -SKITDWTDRSTAVLFGD---GAGAVVMG-----EVAEGRGII-SY 198
QY 134 TIDMSQSGIFYGVYQVDGVVK-----RMKQAESEPKVIATGGLAKLIGTSETI 234
Db 199 EMGSDSGGKYLIDRETGKLMNGREVEKFAVRIMGDASTRVVEKAGLUS-----SEDI 252
QY 235 DV 236
Db 253 DL 254

RESULT 11

US-08-804-227C-8
; Sequence 3, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN USA
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784

Db 431 GVEDIVLVDNALVIGIETT-----GGVMT-----LIERNTAIP-----TK 467
QY 176 PQVGVGTNTIDSMOS--GIFGVYSQV-----DGVVVKRMAQAESEPKVIAT----- 220
Db 468 KSQIFST-ADVQKAVRIQVYEGERAMVYKDNLLGNFELSDIRAPRGVPOQIEVTFALDA 526
QY 221 GGLAKLIGTSET 233
Db 527 NGILTVSATDKDT 539

RESULT 14
US-08-430-925A-4
; Sequence 4, Application US/08430925A
; Patent No. 5763252
; GENERAL INFORMATION:
; APPLICANT: Skadsen, Ronald W
; APPLICANT: Tibbot, Brian K
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430.925A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93171
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 877 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-430-925A-4

Query Match 5.8%; Score 74.5; DB 1; Length 877;
Best Local Similarity 21.6%; Pred. No. 15;
Matches 57; Conservative 36; Mismatches 92; Indels 79; Gaps 13;

QY 4 VIDGNTNTVLGVYQDETFLVHHRLATSRQKTEDEYAMTVRSLEFHA-GLOFQDIDGIVI 62
Db 118 VLSPAGSDLVLTVAHP-----FRFTVSRSTGD-----TLFDTAPGLVFRD----- 159
QY 63 SSVVPPMFMFLEQMKCKYFHTPMI-----IGPGIKTGLNKKYDNP-----KEVGA 108
Db 160 -----KYLEVTSALPAGRASLYGLGEHTKSFRLRHNDSETLWNADIGA 203
QY 109 DRIVNAVAALYGYPAIVDF---GTATTCLINEKQ---YAGGVIAPIAGMISTEALY 162
Db 204 SYV-----DVLNLSHPYMDVRAPGTAHGVLNLSNGMDVLYGGSYVYTKYVIGVLDYF 258
QY 163 HRASKLPRIETAKPKQVGTNTIDSMQSGIF---YGY--VSQVDGVVVKRMAQAESEPK 216
Db 259 FFAGNPFLAVDQVTLIARPAHPYNSFGHQCRGYLNVSDLERVARY-AKARIPLE 317
QY 217 VIATGGLAKLIGTSETIDVDSF 240

Db 318 VMWT-----DIDYMDGF 329

RESULT 15
US-08-466-390-2
; Sequence 2, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466.390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-390-2

Query Match 5.7%; Score 73.5; DB 1; Length 639;
Best Local Similarity 31.3%; Pred. No. 11;
Matches 21; Conservative 15; Mismatches 26; Indels 5; Gaps 2;

QY 157 STEALYHRASKLPR-----TEIAKPKQVVGTVGTNTIDSMQSGIFYGYVSQVDGVVVKRMAQAE 212
Db 173 RADALLKAKEELEKMKSVIENAKKKEVAGAKPHITAAGKLNMIYDLDNVVKVQA-AQ 231
QY 213 SEPKVIA 219
Db 232 SEAKVVS 238

Search completed: June 24, 2003, 22:10:41
Job time : 11.3055 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.1159 Seconds
(without alignments)
1947.059 Million cell updates/sec

Title: US-09-813-453A-47
Perfect score: 1293
Sequence: 1 MILVIDGNTVTLGVYQDE.....DVIDSFLTLKGLIYKKNV 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 417779 seqs, 108206813 residues
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap:*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap:*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap:*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap:*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap:*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap:*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap:*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap:*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap:*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap:*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap:*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap:*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap:*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1293	100.0	254	9	US-09-813-453A-47
2	987	76.3	258	9	US-09-813-453A-2
3	975	75.4	258	9	US-09-813-453A-49
4	960	74.2	262	9	US-09-813-453A-45
5	862	66.7	233	9	US-09-813-453A-17
6	814.5	63.0	256	9	US-09-813-453A-55
7	771	59.6	255	9	US-09-813-453A-7
8	662	51.2	265	9	US-09-813-453A-4
9	648.5	50.2	250	9	US-09-813-453A-3
10	598	46.2	260	9	US-09-813-453A-51
11	528	40.8	229	9	US-09-813-453A-6
12	527	40.8	219	9	US-09-813-453A-57
13	487.5	37.7	272	9	US-09-813-453A-5
14	483.5	37.4	272	9	US-09-712-363-276
15	437.5	33.8	262	9	US-09-813-453A-8
16	389.5	30.1	246	9	US-09-813-453A-9
17	330.5	25.6	212	9	US-09-813-453A-59
18	322	24.9	257	9	US-09-813-453A-53
19	320.5	24.8	273	9	US-09-813-453A-10

20	275.5	21.3	262	9	US-09-813-453A-11	Sequence 11, Appl
21	197.5	15.3	244	9	US-09-813-453A-41	Sequence 41, Appl
22	179.5	13.9	241	9	US-09-813-453A-63	Sequence 63, Appl
23	138.5	10.7	229	9	US-09-813-453A-12	Sequence 12, Appl
24	137.5	10.6	257	9	US-09-813-453A-13	Sequence 13, Appl
25	137	10.6	460	9	US-09-813-453A-39	Sequence 39, Appl
26	137	10.6	592	9	US-09-813-453A-22	Sequence 22, Appl
27	136	10.5	592	9	US-09-813-453A-43	Sequence 43, Appl
28	131.5	10.2	267	9	US-09-813-453A-15	Sequence 15, Appl
29	121	9.4	248	9	US-09-813-453A-20	Sequence 20, Appl
30	117.5	9.1	249	9	US-09-813-453A-70	Sequence 70, Appl
31	114	8.8	249	9	US-09-813-453A-61	Sequence 61, Appl
32	110	8.5	209	9	US-09-813-453A-21	Sequence 21, Appl
33	107.5	8.3	242	9	US-09-813-453A-65	Sequence 65, Appl
34	91	7.0	223	9	US-09-895-913A-74	Sequence 74, Appl
35	91	7.0	223	9	US-09-813-453A-14	Sequence 14, Appl
36	91	7.0	223	9	US-09-813-453A-67	Sequence 67, Appl
37	84	6.5	225	9	US-09-738-626-4158	Sequence 4158, Ap
38	80.5	6.2	958	9	US-09-953-280-43	Sequence 43, Appl
39	80	6.2	343	9	US-10-060-432-28	Sequence 28, Appl
40	80	6.2	343	10	US-09-905-173-28	Sequence 28, Appl
41	79.5	6.1	518	9	US-09-738-626-4769	Sequence 4769, Ap
42	79	6.1	2492	10	US-09-991-258-3	Sequence 3, Appl
43	79	6.1	4834	9	US-10-097-534-27	Sequence 27, Appl
44	78.5	6.1	317	9	US-09-533-029-4	Sequence 4, Appl
45	78.5	6.1	317	9	US-10-286-264-114	Sequence 114, App

ALIGNMENTS

RESULT 1

US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1

; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813.453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 47

; LENGTH: 254

; TYPE: PFT

; ORGANISM: Bacillus halodurans

US-09-813-453A-47

Query Match 100.0%; Score 1293; DB 9; Length 254;

Best Local Similarity 100.0%; Pred. No. 2.8e-122;

Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MILVIDGNTVTLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLSFDHAGLQFQIDGI 60

Db 1 MILVIDGNTVTLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLSFDHAGLQFQIDGI 60

Qy 61 VISSVPPMMFSLQMKCKYFHTPTMIIGPGIKGLNKKYDNPKEVGADRVLNVAATEL 120

Db 61 VISSVPPMMFSLQMKCKYFHTPTMIIGPGIKGLNKKYDNPKEVGADRVLNVAATEL 120

Qy 121 YGPAIVVDGTTATTCYCLINEKKQYAGGVAPGMISTEALYHRASKLPRIEIAKPKQV 180

Db 121 YGPAIVVDGTTATTCYCLINEKKQYAGGVAPGMISTEALYHRASKLPRIEIAKPKQV 180

Qy 181 GTNTIDSMQSGIFGVYVSQVDGVVVKRMAQAESPKVIATGGLAKLIGTSETIDVIDSF 240

Db 181 GTNTIDSMQSGIFGVYVSQVDGVVVKRMAQAESPKVIATGGLAKLIGTSETIDVIDSF 240

RESULT 3
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Vocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGS-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77

```
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55

Query Match      63.0%; Score 814.5; DB 9; Length 256;
Best Local Similarity 60.2%; Pred. No. 4.6e-74;
Matches 153; Conservative 45; Mismatches 55; Indels 1; Gaps 1;

QY 1 MLVIDGNTNTVLGVYQDETIVHWRLATSRQKTEDEYAMTVRSIFDHAGLQFODIDGI 60
Db 1 MLLVFDVGNNTNVLGIYKGDKNVWRKTRDKREYSDGYLILSNLFDNWNISDIDDV 60

QY 61 VISSVPPMFSLEOMCKKYYFHVTPMIIIGPGIKTGLNPKYDNPKEVGADRVNVAIAEL 120
Db 61 IISVVPPMHSLENFCIKYCKKQPLIVGPGIKTGLNPKYDNPKEVGADRVNVAIAEL 120

QY 121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRSKLPRIEIAKPKQV 180
Db 121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRSKLPRIEIAKPKQV 180

QY 181 GTNTIDSMQSGIFYGVYQVSDGVVVKRMAQAE-SEPKVIATGGLAKLIGTSETIDVIDS 239
Db 181 CKSTVSAMQSGIYGVGLVDKLIISMKKELNCDVVKVATGGLAKLIASETKSIDYVDG 240

QY 240 FLTLKGLQLIYKKN 253
Db 241 FLTLGLRLIYEKN 254

RESULT 7
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-17

Query Match      66.7%; Score 862; DB 9; Length 233;
Best Local Similarity 73.1%; Pred. No. 6.5e-79;
Matches 158; Conservative 27; Mismatches 31; Indels 0; Gaps 0;

QY 1 MLVIDGNTNTVLGVYQDETIVHWRLATSRQKTEDEYAMTVRSIFDHAGLQFODIDGI 60
Db 1 MLLVIDGNTNTVLGVYHKGLEYHWRIETSRKHTDEFGMLRSLFDHSLMFEQIDGI 60

QY 61 VISSVPPMFSLEOMCKKYYFHVTPMIIIGPGIKTGLNPKYDNPKEVGADRVNVAIAEL 120
Db 61 IISVVPPMFALEMRCTKYFHEIPQIVGPGMKTGLNPKYDNPKEVGADRVNVAIAEL 120

QY 121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRSKLPRIEIAKPKQV 180
Db 121 YGNPLIVDFGTATTYCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDNI 180

QY 181 GTNTIDSMQSGIFYGVYQVSDGVVVKRMAQAESEPK 216
Db 181 GKNTVSAMQSGILFYGVQVEGIVKRMKQAKQDPR 216
```

```
RESULT 6
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
```

```
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55

Query Match      63.0%; Score 814.5; DB 9; Length 256;
Best Local Similarity 60.2%; Pred. No. 4.6e-74;
Matches 153; Conservative 45; Mismatches 55; Indels 1; Gaps 1;

QY 1 MLVIDGNTNTVLGVYQDETIVHWRLATSRQKTEDEYAMTVRSIFDHAGLQFODIDGI 60
Db 1 MLLVFDVGNNTNVLGIYKGDKNVWRKTRDKREYSDGYLILSNLFDNWNISDIDDV 60

QY 61 VISSVPPMFSLEOMCKKYYFHVTPMIIIGPGIKTGLNPKYDNPKEVGADRVNVAIAEL 120
Db 61 IISVVPPMHSLENFCIKYCKKQPLIVGPGIKTGLNPKYDNPKEVGADRVNVAIAEL 120

QY 121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRSKLPRIEIAKPKQV 180
Db 121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRSKLPRIEIAKPKQV 180

QY 181 GTNTIDSMQSGIFYGVYQVSDGVVVKRMAQAE-SEPKVIATGGLAKLIGTSETIDVIDS 239
Db 181 CKSTVSAMQSGIYGVGLVDKLIISMKKELNCDVVKVATGGLAKLIASETKSIDYVDG 240

QY 240 FLTLKGLQLIYKKN 253
Db 241 FLTLGLRLIYEKN 254

RESULT 7
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match      59.6%; Score 771; DB 9; Length 255;
Best Local Similarity 55.7%; Pred. No. 1.1e-69;
Matches 141; Conservative 52; Mismatches 60; Indels 0; Gaps 0;

QY 1 MLVIDGNTNTVLGVYQDETIVHWRLATSRQKTEDEYAMTVRSIFDHAGLQFODIDGI 60
Db 1 MLLVIDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGLQIRAV 60

QY 61 VISSVPPMFSLEOMCKKYYFHVTPMIIIGPGIKTGLNPKYDNPKEVGADRVNVAIAEL 120
Db 61 IISVVPPMPLTVGLERLSLGYGMRPLVVGPGIKTGNPQYDNPKEVGADRVNVAIAEL 120

QY 121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRSKLPRIEIAKPKQV 180
Db 121 YRTSLIIVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRSKLPRIEIAKPKQV 180

QY 181 GTNTIDSMQSGIFYGVYQVSDGVVVKRMAQAESEPKVIATGGLAKLIGTSETIDVIDS 240
Db 181 ARNTVNSMQAGIYGVYGLVDVDEIVTRMKAESKADPRVATGGLASLAPIESKTIAVEY 240
```


RESULT 11
US-09-813-453A-6
Query Match 40.8%; Score 528; DB 9; Length 258;
Best Local Similarity 41.3%; Pred. No. 3.4e-45;
Matches 105; Conservative 55; Mismatches 92; Indels 2; Gaps 2;
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 258
TYPE: PR1
ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

Query Match 40.8%; Score 528; DB 9; Length 258;
Best Local Similarity 41.3%; Pred. No. 3.4e-45;
Matches 105; Conservative 55; Mismatches 92; Indels 2; Gaps 2;
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 258
TYPE: PR1
ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

RESULT 12
US-09-813-453A-57
Query Match 37.7%; Score 487.5; DB 9; Length 272;
Best Local Similarity 37.9%; Pred. No. 4.4e-41;
Matches 100; Conservative 64; Mismatches 87; Indels 13; Gaps 4;
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 57
LENGTH: 219
TYPE: PR1
ORGANISM: Dehalococcoides ethenogenes

US-09-813-453A-57

Query Match 40.8%; Score 527; DB 9; Length 219;
Best Local Similarity 43.5%; Pred. No. 3.4e-45;
Matches 91; Conservative 57; Mismatches 61; Indels 0; Gaps 0;
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 272
TYPE: PR1
ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

RESULT 13
US-09-813-453A-5
Query Match 37.7%; Score 487.5; DB 9; Length 272;
Best Local Similarity 37.9%; Pred. No. 4.4e-41;
Matches 100; Conservative 64; Mismatches 87; Indels 13; Gaps 4;
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 272
TYPE: PR1
ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

Query Match 37.7%; Score 487.5; DB 9; Length 272;
Best Local Similarity 37.9%; Pred. No. 4.4e-41;
Matches 100; Conservative 64; Mismatches 87; Indels 13; Gaps 4;
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 272
TYPE: PR1
ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

RESULT 14

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.10944 seconds
(without alignments)
2061.866 Million cell updates/sec

Title: US-09-813-453A-47
Perfect score: 1293
Sequence: 1 MILVIDGNTNTVLGVYQDE.....DVIDSFLTLKGLIYKKNV 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	862	66.7	233	1	YACB_BACSU	P37564 bacillus su
2	131.5	10.2	267	1	BAF_BORPE	Q45338 bordetella
3	101.5	7.8	256	1	H161_PSEAE	Q9hu44 pseudomonas
4	97.5	7.5	425	1	GSA_CLOPE	Q9huc8 clostridium
5	95	7.3	304	1	HEM3_AQUAE	O66621 aquifex aeo
6	93.5	7.2	243	1	Y004_METJA	Q60315 methanococc
7	93.5	7.2	515	1	LEUL_BACST	Q8185 bacillus st
8	92	7.1	243	1	HIS4_BRUME	Q8ye36 brucella me
9	89.5	6.9	482	1	THIL_ECO57	Q8xe74 escherichia
10	88.5	6.8	438	1	FUMC_SULSO	P39461 sulfolobus
11	86	6.7	357	1	BUK_THETN	Q8r832 thermoanaer
12	86	6.7	507	1	FUMH_RAT	P14408 rattus norv
13	85.5	6.6	482	1	THIL_ECOLI	P77718 escherichia
14	85.5	6.6	482	1	THIL_SALTI	Q828x1 salmonella
15	85.5	6.6	482	1	THIL_SALTY	P59113 salmonella
16	84.5	6.5	425	1	IUCD_ECOLI	P11295 escherichia
17	84	6.5	616	1	ILVD_YERPE	Q8zab3 versinia pe
18	83.5	6.5	703	1	PURL_SULTO	Q970v6 sulfolobus
19	82	6.3	237	1	HIS4_THET	Q9rpq5 thermoanaer
20	82	6.3	240	1	HIS4_LISMO	Q9y94 listeria mo
21	82	6.3	293	1	ISPE_LISIN	Q8yael listeria mo
22	82	6.3	2485	1	POLN_EEVV3	P36327 venezuelan
23	82	6.3	2492	1	POLN_EEVVP	P36328 venezuelan
24	82	6.3	3390	1	POLG_DEN3	P27915 d genome po
25	81.5	6.3	247	1	HIS4_RHME	Q92tb2 rhizobium m
26	81.5	6.3	253	1	TPIS_COPFA	P21820 coptis japo
27	81.5	6.3	292	1	KSGA_BACSU	P37468 bacillus su
28	81.5	6.3	448	1	TRME_NEIMA	Q9jwb7 neisseria m
29	81.5	6.3	448	1	TRME_NEIMA	Q9jxl4 neisseria m
30	81.5	6.3	451	1	DPD2_CAEEL	Q19366 caenorhabdi
31	81.5	6.3	1220	1	DDX8_HUMAN	Q14562 homo sapien
32	80	6.2	239	1	HIS4_CLOAB	Q97kh9 clostridium
33	80	6.2	376	1	ARGD_AQUAE	O66442 aquifex aeo

RESULT 1

ID	YACB_BACSU	STANDARD;	PRT;	233 AA.
AC	P37564;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical protein yacB.			
GN	YACB.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
EX	MEDLINE=96051385; PubMed=7584024;			
RA	Ogasawara N., Nakai S., Yoshikawa H.;			
RT	"Systematic sequencing of the 180 kilobase region of the Bacillus			
RL	subtilis chromosome containing the replication origin.";			
RL	DNA Res. 1:1-14(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
EX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruchi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,			
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,			
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus			
RT	subtilis.";			
RL	Nature 390:249-256(1997).			
CC	!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.			

O26819 methanobact
Q58595 methanococc
O67428 aquifex aeo
Q05029 saccharomyc
Q9uhc6 homo sapien
Q29116 sus scrofa
Q95242 sus scrofa
Q27571 drosophila
P10177 e kgk/kdpg
Q92e87 listeria in
Q92f77 listeria in
P50310 cricetus

ALIGNMENTS

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D26185; BAA05305.1; -
 DR EMBL; Z99104; CAB11846.1; -
 DR Subtilist; BG10133; YacB.
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc.factor; 1.
 DR TIGRFAMs; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

 Query Match 66.7%; Score 862; DB 1; Length 233;
 Best Local Similarity 73.1%; Pred. No. 6.7e-33;
 Matches 158; Conservative 27; Mismatches 31; Indels 0; Gaps 0;

 QY 1 MLVLVDGNTNTVLGVYQDQETLVHWRHLATSRKTEDEYAMTVRSFLDHAGLQFDIDGI 60
 DB 1 MLVLVDGNTNTVLGVYHDKLEYHWRIFTSRHKTEDEFGMLRSLFDHSLMFEQIDGI 60
 QY 61 VISSVPPMMFSLQMCCKYFHVPTMIIGPKITGLNPKYDNKEVGDRIVNAVAIEL 120
 DB 61 IISVVPPIMFALERMCTKYFHIEPQIVGPMGTGLNPKYDNKEVGDRIVNAVAIHL 120
 QY 121 YGPAIVVDGTATTTCLINEKQYAGGVITAPGIMISTEALYHRASKLPRIETAKPKQV 180
 DB 121 YGNPLIVVDGTATTTCYIDENQYMGGAIPAGTITSTEALYSRAAKLPRIETRPDNI 180
 QY 181 GTNTIDSMQSGIFGVYQVDGVYVVKRKAQAESEPK 216
 DB 181 GKNTVSAMQSGILFVGVQVEGIVKRMKWAQKQDP 216

 RESULT 2
 BAF_BORPE STANDARD; PRT; 267 AA.
 AC Q45338; Q45373;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Bvg accessory factor.
 GN BAF.
 OS Bordetella pertussis.
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 Bordetella.
 NCBI_TaxID=520;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BP504;
 RX MEDLINE=95325323; PubMed=7601846;
 RA Deshaizer D., Wood G.E., Friedman R.L.;
 RT "Identification of a Bordetella pertussis regulatory factor required
 RT for transcription of the pertussis toxin operon in Escherichia
 RT coli.";
 RL J. Bacteriol. 177:3801-3807(1995).
 RN [2]
 RP SEQUENCE OF 1-38 FROM N.A.
 RC STRAIN-BP504;
 RA Wood G.E., Friedman R.L.;
 RT "Identification of a bira homolog in Bordetella pertussis.";
 RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 239-267 FROM N.A.
 RC STRAIN-BP536;
 RX MEDLINE=96419162; PubMed=8821935;
 RA Allen A.G., Maskell D.J.;
 RT "The identification, cloning and mutagenesis of a genetic locus

required for lipopolysaccharide biosynthesis in Bordetella
 pertussis.";
 Mol. Microbiol. 19:37-52(1996).
 CC -I- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
 CC A BVAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
 CC RNA POLYMERASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12020; AAA75361.1; -
 DR EMBL; AF016461; AAC68834.1; -
 DR EMBL; X90711; CAA62242.1; -
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc.factor; 1.
 KW Transcription regulation; Activator.
 SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

 Query Match 10.2%; Score 131.5; DB 1; Length 267;
 Best Local Similarity 24.1%; Pred. No. 0.0011;
 Matches 63; Conservative 43; Mismatches 98; Indels 57; Gaps 12;

 QY 1 MLVLVDGNTNTVLGVYQDQETLVHWRHLATSRKTEDEYAMTV 42
 DB 1 MLVLVDGNTNTVLGVYHDKLEYHWRIFTSRHKTEDEFGMLRSLFDHSLMFEQIDGI 59
 QY 43 VRSFLDHAGLQFDIDGI VISSVPPMMFSLQMCCKYFHVPTMIIGPKITGLNPKYDN 102
 DB 60 V-----AGL-----ARGEAIAATLRAGGCDI-----RWLRAOPLAM--GLRNG-----YRN 98
 QY 103 PREVGADR-----IVNAVAATIELYCPAIVVDGTATTTCLINEKQYAGGVITAPGIMISTE 159
 DB 99 PDQLGADRWACMVGVILAROPSVHPPLLVASFGTATTLDITGPDNVFPFGLILPQPMARG 158
 QY 160 ALYHRASKLPRIETAKPKQVGVNTIDSMQ-----STGIFYGVYQVDGVYVVKRKAQAE---S 213
 DB 159 ALAYGTAHLPLAD-----GLVADYPTDTHQAIASGI-----AAAQAGAIVRQWLGRQRYGQ 210
 QY 214 EPKVIATGGGLAKLIGTSETI 234
 DB 211 APEIYVAGGWPEVRQEAERL 231

 RESULT 3
 HI61_PSEAE STANDARD; PRT; 256 AA.
 ID HI61_PSEAE
 AC Q9HU44;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Imidazole glycerol phosphate synthase subunit hisF1 (EC 4.1.3.-) (IGP
 DE synthase cyclase subunit) (IGP synthase subunit hisF1) (ImGP synthase
 DE subunit hisF1) (IGPS subunit hisF1).
 GN HISF1 OR PA5140.
 OS Pseudomonas aeruginosa.
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 NCBI_TaxID=287;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;

RESULT 5

```
ID   HEM3_AQUAE      STANDARD;      PRT;      304 AA.
AC   O66621;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DE   Porphobilinogen deaminase (EC 4.3.1.8) (PBG) (Hydroxymethylbilane
DE   synthase) (HMBs) (Pre-uroporphyrinogen synthase).
GN   HEMC OR AQ.263.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC   Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA   "The complete genome of the hyperthermophilic bacterium Aquifex
RA   aeolicus."
RL   Nature 392:353-358(1998).
CC   -1- FUNCTION: TETRAPOLYMERIZATION OF THE MONOPYRROLE PBG INTO THE
CC   HYDROXYMETHYLBILANE PREUROPORPHYRINOGEN IN SEVERAL DISCRETE STEPS.
CC   -1- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O -
CC   hydroxymethylbilane + 4 NH(3).
CC   -1- COFACTOR: COVALENTLY BINDS A DIPPYROMETHANE COFACTOR TO WHICH
CC   THE PORPHOBILINOGEN SUBUNITS ARE ADDED (BY SIMILARITY).
CC   -1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; fourth step.
CC   -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -1- SIMILARITY: BELONGS TO THE HMBs FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; A800681; AAC06580.1; ALT_INT.
DR   HSSP; P06983; 2YPN.
DR   InterPro; IPR000860; Porphobil_deam.
DR   Pfam; PF01379; Porphobil_deam; 1.
DR   PRINTS; PR00151; PORPHBDNMASE.
DR   ProDom; PD002745; Porphobil_deam; 1.
DR   TIGRfams; TIGR00212; hmc; 1.
DR   PROSITE; PS00533; PORPHOBILINOGEN_DEAM; 1.
DR   Porphyrin biosynthesis; Lyase; Complete proteome.
DR   BINDING 236 236 PYRROMETHANE COFACTOR (BY SIMILARITY).
SQ   SEQUENCE 304 AA; 34030 MW; FE9C210FA81DEB6B CRC64;

Query Match 7.3%; Score 95; DB 1; Length 304;
Best Local Similarity 19.4%; Pred. No. 1.1;
Matches 48; Conservative 47; Mismatches 87; Indels 66; Gaps 9;

QY 5 IDVGNNTVLGVYQ-----DETILVHHWRATSRQ---TEDEVAMTVRSFLDHAGLQFDI 57
DB 1 IIGTRKSKALWQAVYVDVFEKHWGVEVELVKITTTGDKITDVPYLAIGKGLFVKRI 62
QY 58 DGVISVPPPMFSLQMKCKYFHTVPMIIGPKTGLNINIKYDNPKEVGADRIYNAVA 117
DB 63 EKALLBGSIDLAVHSLKDV-----PWIPLKGLKGAITKRENPYDLISR-----SG 109
QY 118 IELYGPAIVDFGTATTTCYCLINEKQYAGGVIAFGIMISTEALYHRASKLPRIEIAKPK 177
DB 110 KLYELPS-----GSVICTSLRR-----QVQIKRR 136
QY 178 QVVGNTIDSMQSGIFYGVSVQDGVVKKMAQAESEPKVIATGGTGLKLGITSEITDVI 237
DB 137 RDLKVEVLR-----GNVDTRMKLK-EGLYDAVILAYAGV-KRMGYESEITEVL 183
```

```
QY 238 DSFLTLTKG 245
DB 184 EDFIPAVG 191

RESULT 6
Y004_METJA      STANDARD;      PRT;      243 AA.
ID   Y004_METJA
AC   O60315;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0004.
GN   MJ0004.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcales; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA   jannaschii."
RL   Science 273:1058-1073(1996).
CC   -1- SIMILARITY: A.FERMENTANS ACTIVATOR OF (R)-HYDROXYGLUTARYL-COA
CC   DEHYDRATASE (HGDC), TO E.COLI YJIL AND M.JANNASCHII MJ0800.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U67459; AAB97985.1; -.
DR   HSSP; P11568; 1HUX.
DR   TIGR; MJ0004; -.
DR   InterPro; IPR002731; ATPase_BadF.
DR   Pfam; PF01869; BcrAD_BadF; 1.
DR   ProDom; PD006344; ATPase_BadF; 1.
DR   TIGRfams; TIGR00241; CoA_E_activ; 1.
DR   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 243 AA; 26897 MW; 288C4F4190575E53 CRC64;

Query Match 7.2%; Score 93.5; DB 1; Length 243;
Best Local Similarity 22.0%; Pred. No. 1.2;
Matches 57; Conservative 40; Mismatches 79; Indels 83; Gaps 12;

QY 1 MILVIDGNTNVLGVYQDETILVHHWRATSRQKTEDEVAMTVRSFLDHAGLQFDIDGI 60
DB 1 MILGIDVGSTTTKVMLEDSKII--W-----YKIEDIGV 33
QY 61 VTSSVPPPMFSLQMKCKYFHTVPMIIGPKTGLNINIKYDNPKEVGADRIYNAVA 118
DB 34 IEEDILLKMKVEIQ---KY-----PIDKIYATG-----YGRHKVSFADKIYVEIALGK 80
QY 119 -----ELYGPAIVDFGTATTTCYCL-INEKQYAGGVIAFGIMISTEALYHRA---SK 167
DB 81 GANYFFNEADG-----VIDIGGQDTRKLVKIDKNGKVVDVFLSKCAAGTKFLEKALDILK 136
QY 168 LPRIEIAKPK-----VVGNTIDSMQSG-----GIFYGVSVQDGVVKKR 206
```

```
Db 137 IDKNEINKYKSDNTAKISSMCAVFAESEIISLSKKVPKEGILMGVYESIINRVIPMTNR 196
QY 207 MKAAQSEPKVIATGTGLAK 225
    :| | : : : :| | | |
Db 197 LKIQ-----NIVFSGGVAK 210

RESULT 7
LEU1_BACST
ID LEU1_BACST STANDARD; PRT; 515 AA.
AC Q8RL85;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE 2-Isopropylmalate synthase (EC 4.1.3.12) (Alpha-isopropylmalate
  synthase) (Alpha-IPM synthetase).
GN LEUA
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 7954;
RA Porat I., Vinogradov M., Lu C.-D., Vyazmensky M., Abdelal A.T.,
  Chipman D.M., Barak Z.;
RT "Cloning and characterization of acetohydroxy acid synthase from
  Bacillus stearothermophilus.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of
  acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
  3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC -!- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA =
  acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.
CC -!- PATHWAY: Leucine biosynthesis; first step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
  SYNTHASE FAMILY. LEUA 1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: AY083837; AAL93359.1;
CC PROSITE: PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
CC PROSITE: PS00816; AIPM_HOMOCIT_SYNTH_2; FALSE_NEG.
CC Leucine biosynthesis; Lyase.
CC SEQUENCE 515 AA; 57033 MW; C765B253EF4F4FD CRC64;

Query Match 7.2%; Score 93.5; DB 1; Length 515;
Best Local Similarity 20.2%; Pred. No. 2.8;
Matches 53; Conservative 32; Mismatches 96; Indels 81; Gaps 9;

QY 25 HWRLTSRQKTEDEYAMTVRSLEDFHAGLOFQDIDGIVTSS-VVPPMFSLQMKCKYFHV 83
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 96 HLFATSTPHMVHKLMTPEQVIEA-----VEAVKYAKRFFPIVQWSAEDACKRSELPF 149
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
QY 84 TPMITGPGIKTGLN-----IKYDNPKEVG-----ADR 110
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 150 LAKIVAEIKAGASVINIPDTVGVIITPKYGEIFLYLQNNYQNTIENVSLSAHCDDLGMA 209
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
QY 111 IVNVAATIELYCPAIVDFGTATYCLINEKKQYAGGVIAPGIMISTEALYHR----- 164
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 210 VVNSLSAIE-----HGATQVECTINGIGERAGNAALEEIAV---ALHIRKDYQV 256
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
QY 165 -----ASKLPRIETAKPKQVVGTVNTIDSMQSGIIFYGVYQVDGVVVKRMA 209
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 257 ETRLNLQEIKRSTNLVSKLTGVVPPNKAIVGKNAF-AHESGI-----HQDGVLKRTT 309
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
QY 210 QAESPKVIATGTGLAKIGTES 231
```

```
Db 310 YEIISPVLGVPSNSMWLGKNS 331
    :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
RESULT 8
HIS4_BRUME
ID HIS4_BRUME STANDARD; PRT; 243 AA.
AC Q8YE36;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]
  imidazole-4-carboxamide isomerase (EC 5.3.1.16)
DE (Phosphoribosylformimino-5-aminomidazole carboxamide ribotide
  isomerase).
GN HISA OR BMEI2042.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
  Brucellaceae; Brucella.
NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
  Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
  Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
  Selkov E., Elser P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
  Haselkorn R., Kyprides N., Overbeek R.;
RA "The genome sequence of the facultative intracellular pathogen
  Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-5-[(5-
  phosphoribosylamino)methylideneamino]imidazole-4-carboxamide = 5-
  [(5-phospho-1-deoxyribulos-1-ylamino)methylideneamino]-1-(5-
  phosphoribosyl)imidazole-4-carboxamide.
CC -!- PATHWAY: Histidine biosynthesis; fourth step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: AE009637; AAL53223.1;
CC TIGRFAMs: TIGR00007; His_biosynth; 1.
CC Isomerase; Histidine biosynthesis; Complete proteome.
CC SEQUENCE 243 AA; 25589 MW; 1B264AC48AD199A CRC64;

Query Match 7.1%; Score 92; DB 1; Length 243;
Best Local Similarity 23.6%; Pred. No. 1.5;
Matches 52; Conservative 42; Mismatches 74; Indels 52; Gaps 12;

QY 19 DETLVHWRLATSRQKTEDE-----YAMTVRSLEDFHAGLOFQDIDGIVTSSVVPWFSL 73
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 23 DQATYNEDPAAQAKAFEDQGFELHVVLDNGAF--AG---ESVNGTAVEAILKATK---- 74
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
QY 74 EQMKCKYFHVTPMIIGPGIKTGLNPKYDNPKEVGADRIVNAVAATIELYCPAIVDFGTFA 133
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 75 -----NPVQLGGGIRTLAHI--ENWLSRGLRRVILGTAVR---DPAIVWEACKA 119
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
QY 134 TTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIETAKPKQVVGTVNTIDSMQSGIF 193
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 120 F-----PGQVAVGIDAKGGKVAVEG-WAEASRLGVIELAKKFFEGAGV-----AAII 164
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
QY 194 YGVYQVDGVVVKRMK-----AQAESEPKVIATGTGLAKL 226
    | :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| |
Db 165 YTDIDR-DGVLAGINWDSTLALAEAVSIP-VIASGGGLASM 202
```

```

RESULT 9
THII_ECO57
ID THII_ECO57 STANDARD; PRT; 482 AA.
AC QXK74;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thioline biosynthesis protein thii.
GN THII OR 20526 OR ECO5477.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=83334;
RN [1]
SEQUENCE FROM N.A.
STRAIN=O157:H7 / EDL933 / ATCC 700927;
MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
SEQUENCE FROM N.A.
STRAIN=O157:H7 / RMD 0509952;
MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: required for the synthesis of the thiazole moiety (By
similarity).
CC -1- PATHWAY: Thiamine biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE THII FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE005221; AAG54773.1; -.
CC DR EMBL; AP002551; BAB33900.1; -.
CC DR InterPro; IPR004114; THUMP_dom.
CC DR InterPro; IPR003720; THII.
CC DR Pfam; PF02568; Thii; 1.
CC DR Pfam; PF02926; THUMP; 1.
CC DR TIGRFAMs; TIGR00342; Thii; 1.
CC KW Thiamine biosynthesis; Complete proteome.
CC SEQUENCE 482 AA; 55003 MW; 896B83BA0E2A352C CRC64;

Query Match 6.9%; Score 89.5; DB 1; Length 482;
Best Local Similarity 22.8%; Pred. No. 5.4;
Matches 68; Conservative 37; Mismatches 116; Indels 77; Gaps 17;

QY 8 GNTNTVLGVYQDELTH-VHWRLATSRQKTEDEYATVRSLEF-----HAGLQFQIDGI 60
DB 28 GNTNRVLKHY-DETLAVLRHDIENIEVRKADENO-RLTIRDLATRPICGHILEVEDVPFT 85
QY 61 VISSVPPMMFSL-EQMKCKYFHV-----TPMIIGPKITGL---NKKYDN 102
DB 86 DMHDIKALVQVRDQDKGTCFVYRKRGRKHDFFSIDVRYGGGLNQHIESARVKLN 145

103 PK-----EVGADRIYNVAIAIE--LYGYP-----IVVDFGTATT 135
146 PDVTVHLEVEDDRLLIKRGYEGIGGFGPQTQEDVLISGSGFDGVSYSMLMRRCGRVH 205
136 YCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIA--KPKQYVG--TNTIDSMQSG 191
206 YCFEN-----LGGAAHEIGVRQVAHYLWNRFSGSSHRVFAINPEPVVGEILEKIDGQMG 261
192 IFGYYSQVDGVVVKRKAQAESEPKVIATGGLAKLIGTSETIDVIDSFUTLKGLOLI 249
262 V-----ILKRMVMVRAAS--KVAERYGVQALV--TGEALGVYSS-QTLINRLI 304

RESULT 10
FUMC_SULSO
ID FUMC_SULSO STANDARD; PRT; 438 AA.
AC P39451;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fumarate hydratase class II (EC 4.2.1.2) (Fumarase).
GN FUMC OR SSO1077.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OC NCBI_TaxID=2287;
RN [1]
SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
STRAIN=ATCC 49255 / DSM 5833 / MT-4;
MEDLINE=94102412; PubMed=8276121;
RA Colombo S., Grisa M., Tortora P., Vanoni M.;
RT "Molecular cloning, nucleotide sequence and expression of a
Sulfolobus solfataricus gene encoding a class II fumarase.";
RL FEBS Lett. 337:93-98(1994).
RN [2]
ERRATUM.
RX MEDLINE=94164299; PubMed=8119400;
RA Colombo S., Grisa M., Tortora P., Vanoni M.;
RL FEBS Lett. 340:151-153(1994).
RN [3]
SEQUENCE FROM N.A.
STRAIN=ATCC 35092 / DSM 1617 / P2;
MEDLINE=21332226; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awey M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Chauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RN [4]
CHARACTERIZATION.
RP STRAIN=ATCC 35091 / DSM 1616 / P1;
RX MEDLINE=91086914; PubMed=2124611;
RA Puchegger S., Redl B., Stoeffler G.;
RT "Purification and properties of a thermostable fumarate hydratase
from the archaeobacterium Sulfolobus solfataricus.";
RL J. Gen. Microbiol. 136:1537-1541(1990).
CC -1- FUNCTION: MAY FUNCTION AS AN AUXILIARY AEROBIC ENZYME IN THE
CITRIC ACID CYCLE. IT IS ACTIVE AT TEMPERATURES RANGING FROM 40
DEGREES CELSIUS TO 90 DEGREES CELSIUS, WITH A MAXIMUM ACTIVITY AT
85 DEGREES CELSIUS. THE PH OPTIMUM IS PH 8.0.
CC -1- CATALYTIC ACTIVITY: (S)-malate -> fumarate + H(2)O.
CC -1- PATHWAY: Tricarboxylic acid cycle.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: TO OTHER THERMOSTABLE CLASS II FUMARASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X75402; CAA53156.1; -.
EMBL; AE006726; AAK41339.1; ALT_INIT.
PIR; S40448; S40448.
HSSP; P08417; 1YFM.
Interpro; IPR000362; Fumarate_lyase.
Pfam; PF02006; lyase 1; 1.
PRINTS; PR00149; FUMARATELYASE
PROSITE; PS00163; FUMARATE_LYASES; 1.
Lysine, Tricarboxylic acid cycle; Complete proteome.
ACT_SITE 160 160
BINDING 297 297
SEQUENCE 438 AA; 47911 MW; 8D2A9CF37817EA6C CRC64;
POTENTIAL.
CARBOXYL GROUP (POTENTIAL).

Query Match 6.8%; Score 88.5; DB 1; Length 438;
Best Local Similarity 21.7%; Pred. No. 5.8;
Matches 63; Conservative 43; Mismatches 95; Indels 89; Gaps 15;

```

10 T NTVLGYVQDETLLHHWRLATSRKQTDEYAMTVRSLFDHAGLOFQDIDIGIVISSVVPPM   69
    || | :| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
129 TNRLPALQ-----OITSSLNKAEEYKDVIKAGRTHLRDALPVTGLGOELISAYADAFA     180
                                         :|| :|| :|| :|| :|| :|| :|| :|| :||
70 MFSLEQMCKKYFHPTMIIGPIGTGNIKRYDNPKFYGCADRVNAVAAIELYG-PA---      125
    || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
181 QHEHQVMNILEYYVKLPYG-GTATGTGL----NSHPEFOERVEINEIRITCLGEKPANRF     236
                                     :|| :|| :|| :|| :|| :|| :|| :|| :||
126 -----IVDF-----GATTVCYLNE-----KKQYAGGV I         150
            || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
237 RAMRLLTDLLLLSGALRNIAVDLYRLQGDIRLMFSGPLTG--LNETDLTPQEIIAGSSI     293
                                  :|| :|| :|| :|| :|| :|| :|| :|| :||
151 APG-----TWISTE---ALYHR---ASKLPRIETAKPVKGWNTIDSMSGSIFYG        195
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
294 MPGKTNPVTVEATILLISOVVGLDHFANOAFSLMGFEFLSMGIPLVGNVITQV-----N    347
                                           -VDVG---VKRMKQAEESEKPIATGGKLIGTESET  233
196 YVSQ-----FDISE :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
348 FISALEKMRLVIDGHVANVYMKRKYAESPSLIITI--VSPFIYGDYKAT          395

```

RESULT 11

	BUK_THETN	STANDARD;	PRT;	357 AA.
Q	Q8R832;			
15-JUN-2002	(Rel. 41, Created)			
15-JUN-2002	(Rel. 41, Last sequence update)			
15-JUN-2002	(Rel. 41, Last annotation update)			
E	Probable butyrate kinase (EC 2.7.2.7) (BK) (Branched-chain carboxylic acid kinase).			
Q	BUK OR TTE2201.			
Q	Thermoanaerobacter tengcongensis.			
S	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;			
C	Thermoanaerobacteriaceae; Thermoanaerobacter.			
Q	NCBI_TaxID=119072;			
Q	[1]			
Q	SEQUENCE FROM N.A.			
Q	STRAIN-MB4 / JCM 11007;			
C	MEDLINE=21992816; PubMed=11997336;			
A	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,			
A	Chen Y., Xue Y., Xiang Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,			
A	Tan H., Chen R., Wang J., Yu J., Yang H.;			
A	"A complete sequence of T. tengcongensis genome.";			
A	Genome Res. 12:689-700(2002).			

-|- CATALYTIC ACTIVITY: ATP + 2-butanate -> ADP + butanoyl phosphate.
 -|- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 -|- SIMILARITY: BELONGS TO THE ACETOKINASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AE013165; AAM25356.1; -.
 DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
 DR PROSITE; PS01076; ACETATE_KINASE_2; 1.
 KW Transférase; Kinase; Complete proteome.
 SQ SEQUENCE 357 AA; 39217 MW; 05CDDE09A2E7EF3F CRC64;

Query Match 6.7%; Score 86; DB 1; Length 357;
Best Local Similarity 20.8%; Pred. No. 7.4;
Matches 58; Conservative 44; Mismatches 99; Indels 78; Gaps 12;

```

Qy 1 MILVIDGNTNTVLGYQD-----ETLVHWRLATSRQKTEDEYAMTVRSLEF---HAG 51
Db 5 LILVINGSTKVAVERDEPVTETLRHSTTELSKYKRIIDQEFRTQAIDMLKEG 64
Qy 52 LQFODIDGIV-----ISS---VPPPMFSLFQMCKKYFHTVPMIGPCKTGLNIKY 100
Db 65 ISLSQIDAIVRGGLLAPIESGTVIVNEKMLEDLKKAERGEHASNL-----GAIAY 116
Qy 101 DNPKEVGADR-IYNVAAIELYGPAT--VVDFGTATYCLINEK-----KQY- 145
Db 117 TLAKEHNPAYIVDPVVVDELEDVARITGLPETEKQSFHALNQKATARRLASDLGRKYD 176
Qy 146 -----AGSVIAGIMISTEALYHRASKLPRIEIA-----KPKQVVGVTNTIDSMQS 190
Db 177 EVNLLIAHGGGISVGA-----HRKGRVIDVNDALNGEGPPSPERAGGLPVLIDLVLK 228
Qy 191 GIFYGYVSQDGVVVKRMAQAESEPKVIATGGGLAKLGT 229
Db 229 CYSKYTYFE-----ENKKLIGRGGIVAHLGT 255

```

```

RESULT 12
FUMH_RAT      STANDARD;      PRT;      507 AA.
AC      P14408;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Fumarate hydratase, mitochondrial precursor (EC 4.2.1.2) (Fumarase).
FH.
GN      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxId=10116;
[1]
RN      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP      TISSUE=Liver;
RX      MEDLINE=89123346; PubMed=2914923;
RA      Suzuki T., Sato M., Yoshida T., Tuboi S.;
RT      "Rat liver mitochondrial and cytosolic fumaramases with identical amino
RL      acid sequences are encoded from a single gene.";
RL      J. Biol. Chem. 264:2581-2586(1989).
CC      -1- CATALYTIC ACTIVITY: (S)-malate + fumarate + H(2)O.
CC      -1- PATHWAY: Tricarboxylic acid cycle.
CC      -1- SUBUNIT: HOMOTETRAMER.
CC      -1- SUBCELLULAR LOCATION: Mitochondrial and cytoplasmic.
CC      -1- ALTERNATIVE PRODUCTS: 2 isoforms; a mitochondrial form (shown
CC      here) and a cytoplasmic form; are produced by alternative
CC      initiation.
CC      -1- SIMILARITY: TO OTHER THERMOSTABLE CLASS II FUMARASES.
-----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).

```

```

CC EMBL; J04473; AAA41177.1; -
DR PIR; A31424; UFRT.
DR HSP; P08417; IYFM.
DR InterPro; IPR000362; Fumarate_lyase.
DR Pfam; PF00206; lyase_1; 1.
DR PRINTS; PRO0149; FUMARATELYASE.
DR TIGRfams; TIGR00979; fumC_II; 1.
DR PROSITE; PS00163; FUMARATE_LYASES; 1.
KW Lyase; Tricarboxylic acid cycle; Mitochondrion; Transit peptide;
KW Acetylation; Alternative initiation.
FT TRANSIT 1 41
FT CHAIN 42 507
FT CHAIN 42 507
FT CHAIN 42 507
FT INIT_MET 41 41
FT MOD_RES 42 42
FT ACT_SITE 232 232
FT BINDING 368 368
SQ SEQUENCE 507 AA; 54464 MW; 0304255E04C0D7D8 CRC64;

Query Match 6.7%; Score 86; DB 1; Length 507;
Best Local Similarity 26.4%; Pred. No. 11;
Matches 46; Conservative 31; Mismatches 47; Indels 50; Gaps 11;

QY 85 PMII---GPGIKTGLNIKYDNPKEVGADRVNAVAIAELYG-----YPAIVVD--- 129
DQ 134 PLVVMQGTQTNMNV-----NEVISNR-----AIEMLGELGSKKPVHPNDHVNKQ 182
QY 130 -----FGTATYCLINEKKQYAGVIAPGIMISTEALYHRASKLPRIEIAKPKQVGTNT 184
DQ 183 SSNDTPTPTAHIAALEVHQ-----VLLPGLQKLDALSAK-----EFAQVIKIGRTH 233
QY 185 IDSMOSGI---FYGYVSQVDGVVVRKMAQAESEPKV--IATGLAKLIGPSET 233
DQ 234 QDAVPLTLGFEFSYVQVQYAMERKA--AMPRIYELAAAGTA--VGGLNT 282

RESULT 13
THIL_ECOLI
ID THIL_ECOLI STANDARD; PRT; 482 AA.
AC P77718;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thiamine biosynthesis protein thiI.
GN THI OR B0423.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Nanth A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 253-482 FROM N.A.
RC STRAIN-K12;
RA Backstrom A.D.;
RA Thesis (1996), Cornell University, U.S.A.

```

```

CC -I- FUNCTION: REQUIRED FOR THE SYNTHESIS OF THE THIAZOLE MOIETY.
CC -I- PATHWAY: Thiamine biosynthesis.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE THI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000148; AAC73526.1; -
DR EMBL; U82664; AAB40179.1; -
DR EMBL; U34923; -; NOT_ANNOTATED_CDS.
DR EcoGene; EG13273; thiI.
DR InterPro; IPR004114; THUMP_dom.
DR InterPro; IPR003720; ThiI.
DR Pfam; PF02568; ThiI; 1.
DR Pfam; PF02926; THUMP; 1.
DR TIGRfams; TIGR00342; ThiI; 1.
KW Thiamine biosynthesis; Complete proteome.
SQ SEQUENCE 482 AA; 54973 MW; 72AF449459E45762 CRC64;

Query Match 6.6%; Score 85.5; DB 1; Length 482;
Best Local Similarity 22.9%; Pred. No. 11;
Matches 69; Conservative 37; Mismatches 112; Indels 83; Gaps 17;

QY 8 GNTNTVLGVVQDFTL--VHHRLATSRQKTEDEY-----AMT-----VRSLEFDHAGLQFD 56
DQ 28 GNIRNLKHY-DETLAVVRWHDNIEVRADENQRLAIRDALTRIPGTHHILEVEDVPFTD 86
QY 57 IDGIVLSSVPPMFMFSLEQMKKYEYH-----TPMIGPGIKTGL--NIK 99
DQ 87 MHDIFEKALVQ-----YRDQLEGATFCVVRKRRKHFSSIDVRYVGGGLNQHIESARVK 142
QY 100 YDNPK-----EVGADRVNAVAIAE-LYGYP-----IVYDFGT 132
DQ 143 LTNPDVTVHLEVEDDRLLLIKRGYEGIGGFPICGTQEDVLSLISGGFDSGVSSYMLMRGC 202
QY 133 ATTYCLINEKKQYAGVIAPGIMISTEALYHRASKLPRIEIA--KPKQVVG--TNTIDSM 188
DQ 203 RVHYCFNF-----LGGAAHEIGVRQVAHYLWNRFGSSHRVFRVAINFPVYVGEILEKIDG 258
QY 189 QSGIFVGYVSQVDGVVVRKMAQAESEPKVIATGKLAGLIGTSETIDVIDSFTLKLGL 248
DQ 259 QMGV-----ILKRMVRAAS--KVAERYGVQALV--TGEALQVSS--QTLNLRL 303
QY 249 I 249
DQ 304 I 304

RESULT 14
THIL_SALTI
ID THIL_SALTI STANDARD; PRT; 482 AA.
AC Q828X1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thiamine biosynthesis protein thiI.
GN THI OR STY0464.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CT18;
RA MEDLINE-21534947; PubMed-11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Duggan K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

```

```

MEDLINE-97352701; PubMed-92090650;
Webb E., Claas K., Downs D.M.;
"Characterization of ThII, a new gene involved in thiazole
biosynthesis in Salmonella typhimurium.";
J. Bacteriol. 179.4399-4402(1997).
[2]
SEQUENCE FROM N.A.
STRAIN-LT2 / SGSC1412 / ATCC 700720;
MEDLINE-21534948; PubMed-11677609;
McLellan M., Sanderson K.E., Spleth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
"Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
Nature 413:852-856(2001).
[3]
SEQUENCE OF 365-482 FROM N.A.
Metcalfe W.W., Jiang W., Wanner B.L.;
Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: REQUIRED FOR THE SYNTHESIS OF THE THIAZOLE MOIETY.
-!- PATHWAY: Thiamine biosynthesis.
-!- SUBCELLULAR LOCATION: Cytoplasmic.
-!- SIMILARITY: BELONGS TO THE THII FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; U94901; AAB63031.1; -
EMBL; AE008715; AAL19379.1; -
EMBL; U69493; AAB39648.1; -
StyGene; SG10723; thii.
InterPro; IPR004114; THUMP_dom.
InterPro; IPR003720; THII.
Pfam; PF02568; THII; 1.
Pfam; PF02926; THUMP; 1.
TIGRFAMs; TIGR00342; ThII; 1.
Thiamine biosynthesis; Complete proteome.
CONFLICT 365 365 E -> R (IN REF. 3).
SEQUENCE 482 AA; 54831 MW; 41CC9EE3740CE5D2 CRC64;
Query Match 6.6%; Score 85.5; DB 1; Length 482;
Best Local Similarity 22.8%; Pred. No.11;
Matches 68; Conservative 36; Mismatches 117; Indels 77; Gaps 17;
QY 8 GNTNTVLGVYQDETLL--VHHWRLATSRKQTEDEYAMTVRSLEFD-----HAGLQFQDIDGI 60
Ddb 28 GNIRVLKHY-DETLAVVHRWDNIIEVRAKDENO-RLVIRDALTTRIPGIHILLEVEDVPT 85
QY 61 VISSVVPMMFSL-EQMCKKYFHV-----TPMIIGPIKGTGL---NIKYDN 102
Ddb 86 DMDHDFEALAQYRQLECKTCVVRKRGKHEFFSIEFYERYVGGGLNQHIESARVLTN 145
QY 103 PK-----EYGADRIYNAVAIE-LXGYPA-----IVVDFGTATT 135
Ddb 146 PDVTVHLEVEDDRLLIKRGYEGIGGFPPTQEDVLISLGGDFSGVSSYMLMRRRCRVH 205
QY 136 YCLINEKQYAGGVTAQIMISTEALYHRASKLPRIETIA--RPKQVVG--TWTIDSMQSG 191
Ddb 206 YCFEN-----LGGAAHEIGVRQVAHYLWNRFSGSHRVFVAINFEPVVGVEILEKVDGQMG 261
QY 192 IFYGVSVQDVGWVKEMKAQAESEPKVIATGGLAKLIGTESEIDVIDSLFLTKGLQLI 249
Ddb 262 V-----YLKRMVYRAAS--KVAERYGQVALV--TGEALGVQVSS--OTLTNRLI 304

```

Search completed: June 24, 2003, 22:11:48

Job time : 7.10944 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.4787 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453A-47
Perfect score: 1293
Sequence: 1 MILVIDGNTNTVLGVYQDE.....DVIDSFLTLGLLIYKKNV 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1293	100.0	254	2 F83660	hypothetical prote
2	862	66.7	233	2 S66100	conserved hypotet
3	849	65.7	259	2 AF1102	conserved hypotet
4	848	65.6	259	2 AF1464	conserved hypotet
5	772	59.7	273	2 E97293	probable transcrip
6	662	51.2	265	2 T36391	hypothetical prote
7	598	46.2	261	2 B87489	transcription acti
8	494.5	38.2	274	2 H86937	conserved hypotet
9	487.5	37.7	272	2 A70955	hypothetical prote
10	437.5	33.8	262	2 E75516	conserved hypotet
11	389.5	30.1	246	2 D72320	conserved hypotet
12	320.5	24.8	273	2 D71326	conserved hypotet
13	275.5	21.3	262	2 F70165	conserved hypotet
14	155	12.0	276	2 A12292	hypothetical prote
15	138.5	10.7	229	2 E70465	hypothetical prote
16	137.5	10.6	257	2 S75559	hypothetical prote
17	137	10.6	592	2 H82031	probable biotin-[a
18	136	10.5	592	2 B81009	Bira protein/Bvg a
19	131.5	10.2	267	2 T40327	baf protein - Bord
20	121	9.4	248	2 H83111	hypothetical prote
21	110	8.5	209	2 H81382	hypothetical prote
22	110	8.5	597	2 B69251	probable electron
23	107.5	8.3	223	2 G71887	hypothetical prote
24	107.5	8.3	242	2 A82637	conserved hypotet
25	101.5	7.8	256	2 H83002	imidazoleglycerol-
26	97.5	7.5	425	2 A99571	conserved hypotet
27	97.5	7.5	425	2 T43861	glutamate-l-semial
28	96.5	7.5	756	2 A85061	probable potassium
29	95	7.3	323	2 A70324	hydroxymethylbilan

30	95	7.3	2493	2 S26372	nonstructural poly
31	95	7.3	2493	2 S72349	nonstructural poly
32	94.5	7.3	313	2 E97221	transcription regu
33	94.5	7.3	346	2 AC1702	weakly proteinases
34	93.5	7.2	243	2 D64300	(R)-2-hydroxygluta
35	92.5	7.2	320	2 C81972	probable 3-oxoacyl
36	92	7.1	243	2 AD3507	1-(5-phosphoribosy
37	91.5	7.1	512	2 A98352	hydantoinase homol
38	91.5	7.1	512	2 AE2930	hydantoinase A (im
39	91	7.0	223	2 F64627	hypothetical prote
40	91	7.0	318	2 AB0476	probable periplasm
41	91	7.0	435	2 F86730	glutathione-disulf
42	90.5	7.0	346	2 AC1331	proteinases homolo
43	90.5	7.0	560	2 AB2043	chaperonin GroEL (
44	90	7.0	2054	2 T32413	probable acetyl-Co
45	89.5	6.9	482	2 E90688	probable oxidoredu

ALIGNMENTS

RESULT 1

F83660
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83660
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83660
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA803805.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0086
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 100.0%; Score 1293; DB 2; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.3e-105;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MILVIDGNTNTVLGVYQDETLVHHWRLATSRQKTEDEVAMTVRSLEFDHAGLQFODIDGI	60
Db	1	MILVIDGNTNTVLGVYQDETLVHHWRLATSRQKTEDEVAMTVRSLEFDHAGLQFODIDGI	60
Qy	51	VISSVPPPMFSLQOMCKKYPFHVTPMIIGPGIKTGLNPKYDNPKEVGADRVNVAALIEL	120
Db	51	VISSVPPPMFSLQOMCKKYPFHVTPMIIGPGIKTGLNPKYDNPKEVGADRVNVAALIEL	120
Qy	121	YGYPALVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPQVV	180
Db	121	YGYPALVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPQVV	180
Qy	131	GTNIDSQSGIFGYGVSVQDVYKRMKAQSEPKVIATGGLAKLIGTESFIDVDSF	240
Db	131	GTNIDSQSGIFGYGVSVQDVYKRMKAQSEPKVIATGGLAKLIGTESFIDVDSF	240
Qy	241	LTLLGLQLIYKKNV	254
Db	241	LTLLGLQLIYKKNV	254

RESULT 2

S66100
conserved hypothetical protein yacB - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C:Accession: S66100; E69740
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994
A:Title: Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome
A:Reference number: S65967; MUID:96051385; PMID:7584024
A:Accession: S66100
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <OGA>
A:CROSS-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:dl005847; PID:g467459
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, December 1993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chikara, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaite, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, K.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69740
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <KUN>
A:CROSS-references: GB:299104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:el182003;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yacB
A:Start codon: TTG
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 56.7%; Score 862; DB 2; Length 233;
Best Local Similarity 73.1%; Pred. No. 6.1e-68;
Matches 158; Conservative 27; Mismatches 31; Indels 0; Gaps 0;

QY 1 MLIVDVGNNTVLGVYQDETLLVHWRATSRQKTEDEYAMTVRSFLFDHAGLQFQIDGI 60
DB 1 MLIVDVGNNTVLGVYHDKLEYHWRATSRQKTEDEYAMTVRSFLFDHAGLQFQIDGI 60
QY 61 VISSVPPPMFSLQCKKCYFHYVTPMIGIKTGLNIDYDNPKEVAGDRIVNAVAIEL 120
DB 61 IISVVPPPMFSLQCKKCYFHYVTPMIGIKTGLNIDYDNPKEVAGDRIVNAVAIEL 120
QY 121 YGPAIVDFGATTTCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIETAKPKQV 180
DB 121 YGNPLIVDFGATTTCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIETAKPKQV 180
QY 181 GTNTIDSMOSGIFGYVQSDGVVKKMAQAASEPKVIATGGAKLIGTSETIDVDSF 240
DB 181 GNTVSSMQAGIFGYVQSDGVVKKMAQAASEPKVIATGGAKLIGTSETIDVDSF 240
QY 241 LTLKGLQLLYKKN 253
DB 241 LTLKGLQLLYRRN 253

RESULT 3
AF1102
Conserved hypothetical protein lmo0221 [imported] - *Listeria monocytogenes* (strain EGD-e)
C:Species: *Listeria monocytogenes*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1102
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ok, C.; Schluter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669.
A:Accession: AF1102
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-259 <GLA>
A:CROSS-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0221
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 65.7%; Score 849; DB 2; Length 259;
Best Local Similarity 60.1%; Pred. No. 9.6e-67;
Matches 152; Conservative 50; Mismatches 51; Indels 0; Gaps 0;

QY 1 MLIVDVGNNTVLGVYQDETLLVHWRATSRQKTEDEYAMTVRSFLFDHAGLQFQIDGI 60
DB 1 MLIVDVGNNTVLGVYQDETLLVHWRATSRQKTEDEYAMTVRSFLFDHAGLQFQIDGI 60
QY 61 VISSVPPPMFSLQCKKCYFHYVTPMIGIKTGLNIDYDNPKEVAGDRIVNAVAIEL 120
DB 61 IISVVPPPMFSLQCKKCYFHYVTPMIGIKTGLNIDYDNPKEVAGDRIVNAVAIEL 120
QY 121 YGPAIVDFGATTTCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIETAKPKQV 180
DB 121 YGNPLIVDFGATTTCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIETAKPKQV 180
QY 181 GTNTIDSMOSGIFGYVQSDGVVKKMAQAASEPKVIATGGAKLIGTSETIDVDSF 240
DB 181 GNTVSSMQAGIFGYVQSDGVVKKMAQAASEPKVIATGGAKLIGTSETIDVDSF 240
QY 241 LTLKGLQLLYKKN 253
DB 241 LTLKGLQLLYRRN 253

RESULT 4
AF1464
Conserved hypothetical protein lin0253 [imported] - *Listeria innocua* (strain Clip1126)
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1464
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ok, C.; Schluter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1464
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <GLA>
A:CROSS-references: GB:AL592022; PIDN:CAC95486.1; PID:g16413682; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin0253
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 65.6%; Score 848; DB 2; Length 259;
Best Local Similarity 59.7%; Pred. No. 1.2e-66;
Matches 151; Conservative 52; Mismatches 50; Indels 0; Gaps 0;

QY 1 MLIVDVGNNTVLGVYQDETLLVHWRATSRQKTEDEYAMTVRSFLFDHAGLQFQIDGI 60
DB 1 MLIVDVGNNTVLGVYQDETLLVHWRATSRQKTEDEYAMTVRSFLFDHAGLQFQIDGI 60
QY 61 VISSVPPPMFSLQCKKCYFHYVTPMIGIKTGLNIDYDNPKEVAGDRIVNAVAIEL 120
DB 61 IISVVPPPMFSLQCKKCYFHYVTPMIGIKTGLNIDYDNPKEVAGDRIVNAVAIEL 120
QY 121 YGPAIVDFGATTTCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIETAKPKQV 180
DB 121 YGNPLIVDFGATTTCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIETAKPKQV 180

H86937
conserved hypothetical protein M0232 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H86937
R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hsiao, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.; Taylor, J.E.; Whitehead, S.; Barrett, B.G.
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H86937
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <STO>
A:Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
C:Genetics:
C:Gene: M0232
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.2%; Score 494.5; DB 2; Length 274;
Best Local Similarity 38.9%; Pred. No. 1.1e-35;
Matches 103; Conservative 59; Mismatches 88; Indels 15; Gaps 4;

QY 1 MILVIDGNTNTVLGYV---QDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQD 56
Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVQVQWRIRTESEVTADELALIIDGLI---GDDSER 57
QY 57 IDGIVISSVPPPMFSLQMKCKYFHVTP-MIIGPGIKTGLNKNYKPNKEVGADRIYNVAV 115
Db 58 LAGAAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVEVGADRIYNVCL 117
QY 116 AALELXGYPAIVVDFGTATTCTCLINEKKQYAGVGIAPGIMISTEALYHRSKLPRIETIAK 175
Db 118 AAYDFRKAIVVDFGSSICDVVSAKGEFGAIAPGVQVSSDAAAARSALRRVELAR 177
QY 176 PKQVGTNTIDSMQSGIFGYVSQVGVVKKMAQAE-----SEPKVIATGGLAKLIGT 228
Db 178 PRSVGKNTVECMQAGVFGAGLVGRLVGRMQRQDVEEFSGDLGNRVAVVATGHTAPLL 237
QY 229 TSETIDVDSFLTLKGLIYKKN 253
Db 238 PELHTVDYDRHLTLGLRLVFN 262

RESULT 9
A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70955
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70955
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-272 <COL>
A:Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113976
A:Experimental source: strain H37RV
C:Genetics:
C:Gene: Rv3600c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 37.7%; Score 487.5; DB 2; Length 272;
Best Local Similarity 37.9%; Pred. No. 4.3e-35;
Matches 100; Conservative 64; Mismatches 87; Indels 13; Gaps 4;

QY 1 MILVIDGNTNTVLGYV---QDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQD 56
Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVQVQWRIRTESEVTADELALIIDGLI---GDDSER 57
QY 57 IDGIVISSVPPPMFSLQMKCKYFHVTP-MIIGPGIKTGLNKNYKPNKEVGADRIYNVAV 115
Db 58 LAGAAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVEVGADRIYNVCL 117
QY 116 AALELXGYPAIVVDFGTATTCTCLINEKKQYAGVGIAPGIMISTEALYHRSKLPRIETIAK 175
Db 118 AAYDFRKAIVVDFGSSICDVVSAKGEFGAIAPGVQVSSDAAAARSALRRVELAR 177
QY 176 PKQVGTNTIDSMQSGIFGYVSQVGVVKKMAQAE-----QAESEPKVIATGGLAKLIGT 230
Db 178 PRSVGKNTVECMQAGVFGAGLVGRLVGRMQRQDVEEFSGDLGNRVAVVATGHTAPLL 237
QY 231 SETIDVIDSFLTLKGLIYKKNV 254
Db 238 LHTVDYDRHLTLGLRLVFN 261

RESULT 10
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain RI)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: E75516
R: White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; M.; Shen, M.; Vanatavean, J.J.; Lam, P.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75516
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <WHI>
A:Cross-references: GB:AE001905; GB:AE005513; NID:g6458144; PIDN:AAF10040.1; PID:g645
A:Experimental source: strain RI
C:Genetics:
C:Gene: DR0461
A:Map position: 1
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 33.8%; Score 437.5; DB 2; Length 262;
Best Local Similarity 35.7%; Pred. No. 9.8e-31;
Matches 89; Conservative 56; Mismatches 97; Indels 7; Gaps 4;

QY 2 ILVIDGNTNTVLGYVQ-QDETLLVHHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
Db 6 LLAVIDGNTTTLGLADASGALHTWRTNREMLPDDLALQLHLGLFTLAGAPIP--RAA 63
QY 61 VLSSVPPPMFSLQMKCKYFHVTPMIIGPGIKTGLNKNYKPNKEVGADRIYNVAVAIEL 120
Db 64 VLSSVAPPVGENYALAKRHFMDIDAFVSAENLPDVTVELDTPGSGVGDRLCNLFQAEKY 123
QY 121 YG---YPAIVVDFGTATTCTCLINEKKQYAGVGIAPGIMISTEALYHRSKLPRIETIAKPK 177
Db 124 LGLDLY-AVVDFGTSFNDVVGRRFLGGILATGAQVSADALFAAKLPRIITLQAP 182
QY 178 QVVGNTIDSMQSGIFGYVSQVGVVKKMAQAESEPKVIATGGLAKLIGTSEFIDVI 237
Db 183 TAIGNKTVHALQSLGVFGVAEMVDGLLRIRRAELPGEAVAVATGFSRTVQGCQIDY 242
QY 238 DSFLTLKGL 246
Db 243 DETLTLRGL 251

RESULT 11
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima

C-Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C-Accession: D72320
R: Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A-Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Thermoplasma acidophilum*
A-Reference number: A72200; MUID:99287316; PMID:10360571
A-Accession: D72320
A-Status: preliminary
A-Molecule type: DNA
A-Residues: 1-246 <AR>
A-Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AD35964.1; PID:g4981417
A-Experimental source: strain MSB8
C:Gene: TM0883
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.1%; Score 389.5; DB 2; Length 246;
Best Local Similarity 35.6%; Pred. No. 1.4e-26;
Matches 90; Conservative 54; Mismatches 88; Indels 21; Gaps 8;

QY 1 MLIVDVGNTVGLVGYODETLVHHWRLATSRKTEDEYATVRSLEPDH-----AGLQFQD 56
DB 1 MLIVDVGNTVGLVGYODETLVHHWRLATSRKTEDEYATVRSLEPDH-----AGLQFQD 56
QY 57 IDGIVSVVPPMFMFSEQCKYFHVTPMTIIGPGIKTGL---NIKVDNPREVGADRVN 113
DB 54 IKGIVASVPTQNTVIERFSQKYFHSPIWV---KANGCVKWNK---NPSEVGADRVN 109
QY 114 AVAAELGYPAIVVDGTTATTCYCLINEKKYAGGVIAPGIMISTEALYHRASKLPRIET 173
DB 110 VVAEVKEYGKGIIDMGATTATVDLV-VNGSEYEGGAILPGFFMVSFLRGTAKLPLEV 168
QY 174 AKPKQVGTNTIDSMQSGIFGYVSQVDGVVVKRKAQAESEPKVIATGGLAKLIGTSET 233
DB 169 KPADFVGVKGTENIRLGVNNGSVYALEGIIRIK-EVYGLPVLVTGGQSKIY-KDMIK 226
QY 234 IDVIDSFTLTKGL 246
DB 227 HEIFDEDLTKGV 239

RESULT 12
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C-Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C-Accession: D71326
R: Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin, they, L.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; MCD Science 281, 375-388, 1998
A-Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A-Reference number: A71250; MUID:98332770; PMID:9665876
A-Accession: D71326
A-Status: preliminary
A-Molecule type: DNA
A-Residues: 1-273 <COL>
A-Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271
A-Experimental source: strain Nichols
C:Gene: TP0431
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 24.8%; Score 320.5; DB 2; Length 273;
Best Local Similarity 31.4%; Pred. No. 1.8e-20;
Matches 80; Conservative 57; Mismatches 103; Indels 15; Gaps 5;

QY 2 ILIVDVGNTVGLVGYODE---TLVHHWRLATSRKTEDEYATVRSLEPDHAGLQFQDI 57
DB 1 MLIVDVGNTVGLVGYODE---TLVHHWRLATSRKTEDEYATVRSLEPDHAGLQFQDI 57

QY 58 DGIIVSVVPPMFMFSEQCKYFHVTPMTIIGPGIKTGLNIDKYNP---KEYGADRVNAV 115
DB 51 RDAFISVVPVLTKTIADAVAGISQVQVWFGPMAYEHLVPRPEVRAEIGTDLVANAV 120
QY 116 AAIELYGYPAIVVDGTTATTCYCLINEKKYAGGVIAPGIMISTEALYHRASKLPRIETAK 175
DB 121 AAYHFRSACVVVDGTTATTCYCLINEKKYAGGVIAPGIMISTEALYHRASKLPRIETAK 180
QY 176 PKQVGTNTIDSMQSGIFGYVSQVDGVVVKRKAQAESE---PKVIATGGLAKLIGTES 231
DB 131 PDSVLGKDTTHAVQAGVVRGTLF-----VIRAMIAQCQELGRCRAAVITGGSLRLESSEV 236
QY 232 ETIDVIDSFTLTKGL 246
DB 237 D-FPPIDAQLTSLGL 250

RESULT 13
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C-Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C-Accession: F70165
R: Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; V Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
A-Authors: Smith, H.O.; Venter, J.C.
A-Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A-Reference number: A70100; MUID:98065943; PMID:9403685
A-Accession: F70165
A-Status: preliminary
A-Molecule type: DNA
A-Residues: 1-262 <KLE>
A-Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g268 A-Experimental source: strain B31

Query Match 21.3%; Score 275.5; DB 2; Length 262;
Best Local Similarity 28.1%; Pred. No. 1.5e-16;
Matches 72; Conservative 58; Mismatches 105; Indels 21; Gaps 5;

QY 3 LIVDVGNTVGLVGYODETLVHHWRLATSRKTEDEYATVRSLEPDHAGLQFQDIDGIVI 62
DB 9 LIIDIGNTSIAPALFKDNQVNLFIKMTNLMRLYDEVYSFFEEFDF-----NVNKVFI 62
QY 53 SSVVPPMFMFSEQCKYFHVTPMTIIGPGIKTGLNIDKYNP---NPKV-----VGADRVNA 114
DB 53 SSVVPPMFMFSEQCKYFHVTPMTIIGPGIKTGLNIDKYNP---NPKV-----VGADRVNA 116
QY 115 VAAIELYGYPAIVVDGTTATTCYCLINEKKYAGGVIAPGIMISTEALYHRASKLPRIET 173
DB 117 VAAIENYSFENVLVVDIGTACTIFAVSRQDGLGILGINSGLPINFNSLDDNAYLLKKPFI 176
QY 174 AKPKQVGTNTIDSMQSGIFGYVSQVDGVVVKRKAQAESEPKVIATGGLAKLIGTES 233
DB 177 STPNLLERTTSGSVNSGLFYQYKLYIEGVYRDIKOMYKKKFNLIITGNADLILSLIEI 236
QY 234 IDVIDSFTLTKGLQLI 249
DB 237 EFIFNIHLTVGEVRIL 252

RESULT 14
AT2292
hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A-Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C-Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C-Accession: AT2292
R: Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriga Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata DNA Res. 8, 205-213, 2001
A-Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12292
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-276 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075595.1; PID:g17133030; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3896

Query Match 12.0%; Score 155; DB 2; Length 276;
Best Local Similarity 22.9%; Pred. No. 5.5e-06;
Matches 60; Conservative 46; Mismatches 116; Indels 40; Gaps 8;
QY 3 LVIDVGNNTVLGVYQDETLLVHWRLL-----ATSRKTEDEYAMTVRSFLDHAGLQFQDI 57
DB 15 LAIEIGNSRLHWALFWGESLEFWDTEYLPESVIOQLNGETKLEVGSEKEIFFTFPL 74
QY 58 D-----GIVISSVPPMFSLEQCKKYPHTVPMIIGPGIKTGLNKKYDNPNKEVGADRI 111
DB 75 PPAPCPPLPLFIASVVPQTV---LWENYLNVRVITLD---QIPLNNIYPT---LGIDRA 124
QY 112 VNAVAIELYGYPAIVVDFGTATYCLINEKKOYAGVIAPGIMISTEALYHRASKLPRI 171
DB 125 LALWAGMSWGFVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQLPQV 184
QY 172 EIAKPKQV---VGTNTIDSMQSGIFYGVYVSOVDGVVVRMKAQAESEPKVIATGGLAKLIG 228
DB 185 EMEAIKSLPRFALNTEAIQSGVITVLIANGMRDFTFEW---LSLFPDGKVAIKGG----- 237
QY 229 TESETIDVIDSFLTKLGLQLIY 250
DB 238 -----DRILLNVLQALY 250

RESULT 15

E70465
hypothetical protein aq_1924 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: E70465
R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70465
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-229 <AQF>
A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07720.1; PID:g2984188; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_1924

Query Match 10.7%; Score 138.5; DB 2; Length 229;
Best Local Similarity 20.9%; Pred. No. 0.00012;
Matches 52; Conservative 54; Mismatches 114; Indels 29; Gaps 7;
QY 3 LVIDVGNNTVLGVYQDETLLVHWRLLATSRKTEDEYAMTVRSFLDHAGLQFQDIDGIVI 62
DB 4 LTVDVGNSSVDIALWEGKKVKDFKL--SHEEFLKEEFPKLAL---GISVKQSFSEKV 57
QY 63 SSVVPPMFSLEQCKKYPHTVPMIIGPGIKTGLNKKYDNPNKEVGADRIVNAVAIELYG 122
DB 58 RGKIPKIKF-----LKKENFPI-----QVDYKTPETGLTDRVALAYSARKFYG 100
QY 123 YPAIVVDFGTATYCLINEKKOYAGVIAPGIMISTEALYHRASKLPRIEIAKPKQVVG 182
DB 101 KNVVVISAGTALVDLVLECK-FKGGFITLGLGKKILSDLAEGIPPEFPEVEIFLGR 159
QY 183 NTIDSMQSGIFYGVYVSOVDGVVVRMKAQAESEPKVIATGGLAKLICTESETIDVIDSFLT 242

Db 160 STRECVLGGAYRESTEFIKSTLKLWRKVPKRKFKVVITGEGKYF-----SKFGIYDPLLV 215
QY 243 LKGLQ-LIY 250
DB 216 HRCMRNLLY 224
Search completed: June 24, 2003, 22:19:15
Job time : 12.4787 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

*OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 28.7514 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453a-47

Perfect score: 1293

Sequence: 1 MILVIDGNTNTVLGVQDE.....DVIDSFLTKGLQLIYKKNV 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_101002.*

```

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1293	100.0	254	AAU91171	Pantothenate kinas
2	987	76.3	258	AAU01243	B. subtilis novel
3	987	76.3	258	AAU91149	Bacillus subtilis
4	975	75.4	258	AAU91172	Pantothenate kinas
5	960	74.2	262	AAU91170	Pantothenate kinas
6	862	66.7	233	AAU91163	Pantothenate kinas
7	849	65.7	239	ABBA7661	Listeria monocytog
8	814.5	63.0	256	AAU91175	Pantothenate kinas
9	771	59.6	255	AAU91154	Geobacter sulfurre
10	662	51.2	265	AAU91151	Streptomyces coeli

11	648.5	50.2	250	23	AAU91150	Clostridium acetob
12	598	46.2	260	23	AAU91173	Pantothenate kinas
13	528	40.8	258	23	AAU91153	Rhodobacter capsul
14	527	40.8	219	23	AAU91176	Pantothenate kinas
15	487.5	37.7	272	23	AAU91152	Mycobacterium tube
16	483.5	37.4	272	22	AAAG81225	Mycobacterium radiop
17	437.5	33.8	246	23	AAU91155	Deinococcus maritim
18	389.5	30.1	246	23	AAU91156	Thermotoga maritim
19	330.5	25.6	212	23	AAU91177	Pantothenate kinas
20	325	24.9	257	23	AAU91174	Pantothenate kinas
21	320.5	24.8	273	23	AAU91157	Treponema pallidum
22	275.5	21.3	262	23	AAU91158	Borrelia burgdorfe
23	197.5	15.3	244	23	AAU91168	Pantothenate kinas
24	179.5	13.9	241	23	AAU91179	Pantothenate kinas
25	138.5	10.7	229	23	AAU91159	Aquifex aeolicus p
26	137.5	10.6	257	23	AAU91160	Synechocystis pant
27	137	10.6	455	20	AAAY38617	Neisseria gonorrhoe
28	137	10.6	455	21	AAAY74908	Neisseria gonorrhoe
29	137	10.6	455	21	AAAY74910	Neisseria meningit
30	137	10.6	460	23	AAU91167	Pantothenate kinas
31	137	10.6	592	20	AAAY38618	Neisseria gonorrhoe
32	137	10.6	592	20	AAAY38616	Neisseria meningit
33	137	10.6	592	21	AAAY74911	Neisseria gonorrhoe
34	137	10.6	592	21	AAAY74913	Neisseria meningit
35	137	10.6	592	23	AAU91166	Pantothenate kinas
36	136	10.5	389	21	AAAY74909	Neisseria meningit
37	136	10.5	592	20	AAAY38615	Neisseria meningit
38	136	10.5	592	21	AAAY74912	Neisseria meningit
39	136	10.5	592	23	AAU91169	Pantothenate kinas
40	131.5	10.2	267	23	AAU91162	Bordetella pertussis
41	121	9.4	248	23	AAU91164	Pantothenate kinas
42	117.5	9.1	249	23	AAU91182	Pantothenate kinas
43	114	8.8	249	23	AAU91178	Pantothenate kinas
44	110	8.5	209	23	AAU91165	Pantothenate kinas
45	107.5	8.3	242	23	AAU91180	Pantothenate kinas

ALIGNMENTS

```

RESULT 1
AAU911171
ID AAU911171 standard; Protein; 254 AA.
XX
AC AAU911171;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #9.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus halodurans.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI: 2002-269358/31.
DR N-PSDB; ABE54192.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein

```

PT with test compound and identifying inhibitor of the Coax protein -
 XX Claim 10; Page 100; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 XX described in the invention.

SQ Sequence 254 AA;
 Query Match 100.0%; Score 1293; DB 23; Length 254;
 Best Local Similarity 100.0%; Pred. No. 3.2e-130;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MILVIDGNTNTVLGYQDETIVHWRLATSRKTEDEYAMTVRSFLFDHAGLQFQDIDGI 60
 1 MILVIDGNTNTVLGYQDETIVHWRLATSRKTEDEYAMTVRSFLFDHAGLQFQDIDGI 60
 61 VISSVPPMMFSLQCKKFFHYTPMIGPKITGLNKKYDNPKEVGADRVNAVAIEL 120
 61 VISSVPPMMFSLQCKKFFHYTPMIGPKITGLNKKYDNPKEVGADRVNAVAIEL 120
 121 YGPAIVVDGTTATTCYCLINEKKQYAGGVAPGIMISTEALYHRASKLPRIEIAKPKQVV 180
 121 YGPAIVVDGTTATTCYCLINEKKQYAGGVAPGIMISTEALYHRASKLPRIEIAKPKQVV 180
 181 GTNTIDSMQSGIFGYVSVQDGVVKKRKAQAESEPKVIATGGLAKLIGTSETIDVDSF 240
 181 GTNTIDSMQSGIFGYVSVQDGVVKKRKAQAESEPKVIATGGLAKLIGTSETIDVDSF 240
 241 LTLKGLQLIYKKNV 254
 241 LTLKGLQLIYKKNV 254

RESULT 2
 AAU01243
 ID AAU01243 standard; Protein; 258 AA.
 XX
 AC AAU01243;
 XX

18-JUL-2001 (first entry)

B. subtilis novel pantothenate kinase encoded by the gene coax.
 Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;
 nutritional supplement; panto-compound; pantoate.

Bacillus subtilis.

WO200121772-A2.

29-MAR-2001.

21-SEP-2000; 2000WO-US25993.

21-SEP-1999; 99US-0400494.

07-JUN-2000; 2000US-0210072.

28-JUL-2000; 2000US-0221836.

24-AUG-2000; 2000US-0227860.

(OMNI-) OMNIGENE BIOPRODUCTS.

Yocum RR, Patterson TA, Hermann T, Pero JG;

WPI; 2001-218644/22.

DR N-PSDB; AAS00984.

XX New recombinant microorganism which overexpress a Bacillus subtilis
 PT pantothenate biosynthetic enzyme, useful for the high yield production
 PT of panto-compounds such as pantothenate and pantoate -
 XX Example 14; Fig 23; 292pp; English.

XX The sequence represents a novel B. subtilis pantothenate kinase (encoded
 CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.
 CC Pantothenate, also known as vitamin B5, is used as a nutritional
 CC supplement in mammals and humans. The invention concerns methods of
 CC producing recombinant microorganisms overexpressing at least one Bacillus
 CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods
 CC of producing them are useful for producing a panto-compound such as
 CC pantothenate or pantoate, which is a nutritional requirement for
 CC livestock and humans. The methods are also useful for the identification
 CC of pantothenate kinase modulators. Panto-compounds are produced at a
 CC significantly higher yield than prior art methods and can be produced
 CC independent of the need to feed precursors which decreases expense.

SQ Sequence 258 AA;

Query Match 76.3%; Score 987; DB 22; Length 258;
 Best Local Similarity 72.3%; Pred. No. 2.3e-97;
 Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

QY 1 MILVIDGNTNTVLGYQDETIVHWRLATSRKTEDEYAMTVRSFLFDHAGLQFQDIDGI 60
 Db 1 LLLVIDGNTNTVLGYHDGKLEYHRIETSRKTEDEFGMILRSUFDHSLMFEQIDGI 60
 QY 61 VISSVPPMMFSLQCKKFFHYTPMIGPKITGLNKKYDNPKEVGADRVNAVAIEL 120
 Db 61 IISVVPPIMFALERMCTKYFHIEPQIVGPKMTGLNKKYDNPKEVGADRVNAVAIEL 120
 QY 121 YGPAIVVDGTTATTCYCLINEKKQYAGGVAPGIMISTEALYHRASKLPRIEIAKPKQVV 180
 Db 121 YGNPLIVVDGTTATTCYCLINEKKQYAGGVAPGIMISTEALYHRASKLPRIEIAKPKQVV 180
 QY 181 GTNTIDSMQSGIFGYVSVQDGVVKKRKAQAESEPKVIATGGLAKLIGTSETIDVDSF 240
 Db 181 GRNTVSAMSGILFGYGVQGVGIVKRMKQAKODLVKVIATGGLAPLIANESDCIDIVDPF 240
 QY 241 LTLKGLQLIYKKNV 253
 Db 241 LTLKGLQLIYKKNV 253

RESULT 3

AAU91149
 ID AAU91149 standard; Protein; 258 AA.

XX

AC AAU91149;

XX

DT 05-JUN-2002 (first entry)

XX

DE Bacillus subtilis pantothenate kinase Coax.

XX

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX Pantothenate kinase modulator; coenzyme A; Bactericidal compound.

OS Bacillus subtilis.

XX

PN WO200216601-A2.

XX

PD 28-FEB-2002.

XX

PF 24-AUG-2001; 2001WO-US26531.

XX

PR 24-AUG-2000; 2000US-227860P.

XX

PR 20-MAR-2001; 2001US-0813453.

XX

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX Yocum RR, Patterson TA;
 PI WPI; 2002-269358/31.
 XX N-PSDB; ABK54191.
 DR Identifying potential antibiotic or antimicrobial agent, comprises
 XX contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein
 XX Claim 8; Page 98-99; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX SQ Sequence 262 AA;
 Query Match 74.2%; Score 960; DB 23; Length 262;
 Best Local Similarity 69.2%; Pred. No. 1.8e-94;
 Matches 175; Conservative 35; Mismatches 43; Indels 0; Gaps 0;
 QY 1 MLIVDVGNTNTVLGVYQDETLLVHWRATSRKTEDEVAMTVRSFLFDHAGLQFQDIDGI 60
 DB 1 MIFVLDGNTNAVGLVFESEGLRQHRMETDRKTEDEVGMVQLLEHGLSFEDVKGI 60
 QY 61 VISSVPPMMFSLQCKKYFHYTPMIGIKTGLNKNYDNPKEVGADRVNAVAIEL 120
 DB 61 IVSSVPPIMFALERMCYFKPLVGVGKTKGLNKNYDNPKEVGADRVNAVAIHL 120
 QY 121 YGPAIVDVGTTATTCYCLINEKKQYAGGVTAIPGIMISTEALYHRASKLPRIETAKPKQV 180
 DB 121 YGSLIIVDVGTTATTCYCLINEKKYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
 QY 181 GTNTIDSMQSGIFGYVQSDGVVKKMAQAESEPKVIATGGTGLAKLIGTSEIDVDSF 240
 DB 181 GKNTVSAMQSGILYGVQGVGIVKRMKEAKQEPKVIATGGTGLAKLISESNVIDVDPF 240
 DB 241 LTLKGLQLYKKN 253
 DB 241 LTLKGLYLYERN 253
 RESULT 6
 AAU91163
 ID AAU91163 standard; Protein: 233 AA.
 XX AC AAU91163;
 XX 05-JUN-2002 (first entry)
 XX Pantothenate kinase (Coax) #1.
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX Bacillus subtilis.
 OS WO200216601-A2.
 XX PN 28-FEB-2002.
 XX PD 28-FEB-2002.
 XX

PF 24-AUG-2001; 2001WO-US26531.
 XX 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA Yocum RR, Patterson TA;
 PI WPI; 2002-269358/31.
 XX N-PSDB; ABK54169.
 DR Identifying potential antibiotic or antimicrobial agent, comprises
 XX contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein
 XX Claim 8; Page 81-82; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX SQ Sequence 233 AA;
 Query Match 66.7%; Score 862; DB 23; Length 233;
 Best Local Similarity 73.1%; Pred. No. 5.1e-84;
 Matches 158; Conservative 27; Mismatches 31; Indels 0; Gaps 0;
 QY 1 MLIVDVGNTNTVLGVYQDETLLVHWRATSRKTEDEVAMTVRSFLFDHAGLQFQDIDGI 60
 DB 1 MLIVDVGNTNTVLGVYHDKLEYHWRATSRKTEDEVGMILRSFLDHSGLMFEQIDGI 60
 QY 61 VISSVPPMMFSLQCKKYFHYTPMIGIKTGLNKNYDNPKEVGADRVNAVAIEL 120
 DB 61 IISVPPIMFALERMCYFKFHEPQIVGPMKTKGLNKNYDNPKEVGADRVNAVAIHL 120
 QY 121 YGPAIVDVGTTATTCYCLINEKKQYAGGVTAIPGIMISTEALYHRASKLPRIETAKPKQV 180
 DB 121 YGNPLIIVDVGTTATTCYCLINEKKQYMGGAIPGIMISTEALYSRAAKLPRIETRPDNI 180
 QY 181 GTNTIDSMQSGIFGYVQSDGVVKKMAQAESEPK 216
 DB 181 GKNTVSAMQSGILFYGVQGVGIVKRMKQAKODPR 216
 RESULT 7
 ABB47661
 ID ABB47661 standard; Protein: 259 AA.
 XX AC ABB47661;
 XX 05-FEB-2002 (first entry)
 XX Listeria monocytogenes protein #365.
 DE Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX Listeria monocytogenes.
 OS WO200177335-A2.
 XX PN 18-OCT-2001.
 XX PD 11-APR-2001; 2001WO-FR01118.
 XX

11-APR-2000; 2000FR-0004629.

(INSP) INST PASTEUR.

Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
Domínguez-Bernal G, Garrido-García P, Tíerrez-Martínez A, Anand A;
Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L;
Perez-Díaz J, Baquero F, García Del Portillo F, Gomez-Lopez N;
Maduenlo E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
Rose M, Voss H;

WPI; 2002-010914/01.

Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment and prevention of *Listeria* and related bacterial infections, and related polypeptides

Claim 6; SEQ ID No 366; 192pp; French.

The present invention relates to the genome sequence of *Listeria monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of it are useful for selecting probes and primers for detecting genes in *L. monocytogenes* and related organisms, and for studying genetic polymorphisms and other genomes. The present sequence is a protein encoded by the genome sequence of the present invention. Proteins expressed from the genome sequence are useful for raising specific antibodies. Identification of *L. monocytogenes* and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of Vitamin B12. The genome sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate *L. monocytogenes*-related diseases. In addition, the genome sequence and proteins encoded by it are useful in pharmaceutical and vaccines compositions for the treatment or prevention of infections by *L. monocytogenes* and related organisms.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 259 AA;

Query Match 65.7%; Score 849; DB 23; Length 259;

Best Local Similarity 60.1%; Pred. NO. 1.5e-82;

Matches 152; Conservative 50; Mismatches 51; Indels 0; Gaps 0;

1 MILVIDGNTNTVLGVYQDETILVHHWRLATSRQKTEDEYAMTVRSFLPDHAGLQFQDIDGI 60

1 MILVIDGNTNTVLGVYQDETILVHHWRLATSRQKTEDEYAMTVRSFLPDHAGLQFQDIDGI 60

61 VISSVPPMPSLEOMCKKYFHVTPMIIIGPGIKTGLNIDYKPNKVEGADRIYNAVAATEL 120

61 VISSVPPMPSLEOMCKKYFHVTPMIIIGPGIKTGLNIDYKPNKVEGADRIYNAVAATEL 120

121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQV 180

121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQV 180

181 GTNIDSQSGIFGYVSDGVVKKRKAQAESEPKVIATGGLAKLTGETSETIDVIDSF 240

181 GTNIDSQSGIFGYVSDGVVKKRKAQAESEPKVIATGGLAKLTGETSETIDVIDSF 240

241 LTLKGLQLIYKKN 253

241 LTLKGLQLIYKKN 253

RESULT 8

AAU91175

ID AAU91175 standard; Protein; 256 AA.

XX AAU91175;

XX

05-JUN-2002 (first entry)

Pantothenate kinase (Coax) #13.

Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.

Clostridium difficile.

WO200216601-A2.

28-FEE-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

N-PSDB; ABK54196.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein

Claim 6; Page 105; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 256 AA;

Query Match 63.0%; Score 814.5; DB 23; Length 256;

Best Local Similarity 60.2%; Pred. No. 7.2e-79;

Matches 153; Conservative 45; Mismatches 55; Indels 1; Gaps 1;

1 MILVIDGNTNTVLGVYQDETILVHHWRLATSRQKTEDEYAMTVRSFLPDHAGLQFQDIDGI 60

1 MLLVFDGNTNVLGIYKDKLVNRYRIKTRDKTSDEYGLISNLFYDNNVNSIDDDV 60

61 VISSVPPMPSLEOMCKKYFHVTPMIIIGPGIKTGLNIDYKPNKVEGADRIYNAVAATEL 120

61 VISSVPPMPSLEOMCKKYFHVTPMIIIGPGIKTGLNIDYKPNKVEGADRIYNAVAATEL 120

121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQV 180

121 YGPAIVVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQV 180

131 GTNIDSQSGIFGYVSDGVVKKRKAQAESEPKVIATGGLAKLTGETSETIDVIDS 239

131 CKSTVSAMQSGIYGYVGLVDKIIISMKKELNCDVKVIATGGLAKLTGETSETIDVIDS 240

240 FLTLKGLQLIYKKN 253

241 FLTLKGLQLIYKKN 254

RESULT 9

AAU91154

ID AAU91154 standard; Protein; 255 AA.

XX

AAU91150 standard; Protein; 250 AA.
AAU91150;
05-JUN-2002 (first entry)
Clostridium acetobutylicum pantothenate kinase Coax.
Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.
Clostridium acetobutylicum.
WO200216601-A2.
28-FEB-2002.
24-AUG-2001; 2001WO-US26531.
24-AUG-2000; 2000US-227860P.
20-MAR-2001; 2001US-0813453.
(OMNI-) OMNIGENE BIOPRODUCTS INC.
Yocum RR, Patterson TA;
WPI; 2002-269358/31.
Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
Claim 10; Page 68-69; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.
Sequence 250 AA;
Query Match 50.2%; Score 648.5; DB 23; Length 250;
Best Local Similarity 53.6%; Pred. No. 4.5e-61;
Matches 126; Conservative 41; Mismatches 57; Indels 11; Gaps 2;
QY 1 MLVIDGNTNTVLGVYODETLVHHWRLATSRQKTEDEYAMTVRSLSFDHAGLQFDIDGI 60
DB 18 VILVDGNTNVLGIYNDTKLTAERLUSTDVLRSADYGVGNLFOODKLDPLVEGV 77
QY 61 VISSVPPMFSLEQCKKYYFHVPMIIGPGIKTGLNIDKYNPKVEGADRVNAAAEI 120
DB 78 IISSVPPNIMYSLEHMRKYPKINPLVVGPGIKTGINKIDNPKVEGADRVNAAAEI 137
QY 121 YGPAIVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEAKPKOV 180
DB 138 YKRSIIIDFGTATTCFAVRENGDYLGAICPGIKVSSEALFEKAAKLPVELIKPAYAI 197
QY 181 GTNTIDSMQSGIFGVYQVSDGVVVKRMAQAESEPK-----VIATGGLAKLI 227
DB 198 CKNTISSQSGIVYRLRQVKYLFKLA---ENLPDGRRTSLVATGGLAKLI 249
RESULT 12
AAU91173
ID AAU91173 standard; Protein; 260 AA.
XX
AC AAU91173;
05-JUN-2002 (first entry)
Pantothenate kinase (Coax) #11.
Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.
Caulobacter crescentus.
WO200216601-A2.
28-FEB-2002.
24-AUG-2001; 2001WO-US26531.
24-AUG-2000; 2000US-227860P.
20-MAR-2001; 2001US-0813453.
(OMNI-) OMNIGENE BIOPRODUCTS INC.
Yocum RR, Patterson TA;
WPI; 2002-269358/31.
N-PSDB; ABK54194.
Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
Claim 10; Page 102-103; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.
Sequence 260 AA;
Query Match 46.2%; Score 598; DB 23; Length 260;
Best Local Similarity 43.9%; Pred. No. 1.3e-55;
Matches 112; Conservative 61; Mismatches 80; Indels 2; Gaps 1;
QY 1 MLVIDGNTNTVLGVYODETLVHHWRLATSRQKTEDEYAMTVRSLSFDHAGLQFDIDGI 60
DB 1 MLLAEQNTNTNFAIHGASVVAQWRSTESTRTADEYVWVLSQGLGFRADIV 60
QY 61 VISSVPPMFSLEQCKKYYFHVPMIIGPGIKTGLNIDKYNPKVEGADRVNAAAEI 120
DB 61 IISSVPPQSIIFNLNLSRRYFNVPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAV 120
QY 121 YGPAIVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEAKPK--KQ 178
DB 121 YPGPLVWDSGTATTFDIVAADGAPEGGIAPGINSQALHEAAKLPRIAIQRPAGNR 180
QY 179 VVGNTIDSMQSGIFGVYQVSDGVVVKRMAQAESEPKVIATGGLAKLIGTSETIDV 238
DB 181 IVGDTVSAVMSQSGVFWGYISLIEGLVARIKAEGERPEMTVIATGVASLFEGATSDIDHD 240
QY 239 SFLTGLQLIYKKN 253
DB 241 SDLTIRGLEIYRN 255
RESULT 13
AAU91153
ID AAU91153 standard; Protein; 258 AA.

XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #11.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
OS Caulobacter crescentus.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54194.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
XX
PS Claim 10; Page 102-103; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.
XX
SQ Sequence 260 AA;
Query Match 46.2%; Score 598; DB 23; Length 260;
Best Local Similarity 43.9%; Pred. No. 1.3e-55;
Matches 112; Conservative 61; Mismatches 80; Indels 2; Gaps 1;
QY 1 MLVIDGNTNTVLGVYODETLVHHWRLATSRQKTEDEYAMTVRSLSFDHAGLQFDIDGI 60
DB 1 MLLAEQNTNTNFAIHGASVVAQWRSTESTRTADEYVWVLSQGLGFRADIV 60
QY 61 VISSVPPMFSLEQCKKYYFHVPMIIGPGIKTGLNIDKYNPKVEGADRVNAAAEI 120
DB 61 IISSVPPQSIIFNLNLSRRYFNVPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAV 120
QY 121 YGPAIVDFGTATTYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEAKPK--KQ 178
DB 121 YPGPLVWDSGTATTFDIVAADGAPEGGIAPGINSQALHEAAKLPRIAIQRPAGNR 180
QY 179 VVGNTIDSMQSGIFGVYQVSDGVVVKRMAQAESEPKVIATGGLAKLIGTSETIDV 238
DB 181 IVGDTVSAVMSQSGVFWGYISLIEGLVARIKAEGERPEMTVIATGVASLFEGATSDIDHD 240
QY 239 SFLTGLQLIYKKN 253
DB 241 SDLTIRGLEIYRN 255
RESULT 13
AAU91153
ID AAU91153 standard; Protein; 258 AA.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 10.6301 seconds
(without alignments)
725.188 Million cell updates/sec

Title: US-09-813-453A-45
Perfect score: 1341
Sequence: 1 MIFVLDVGNFNAVLGVFEFG.....LKGLMYERNALQHEKGE 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pap.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92.5	6.9	361	3	US-09-032-372-12
2	90	6.7	1233	4	Sequence 12, Appl
3	86	6.4	357	4	Sequence 5, Appl
4	86	6.4	560	3	Sequence 826, App
5	86	6.4	1228	4	Sequence 13, Appl
6	86	6.4	1228	4	Sequence 537, App
7	86	6.4	1261	4	Sequence 537, App
8	86	6.4	1261	4	Sequence 538, App
9	85.5	6.4	288	4	Sequence 439-313-538
10	83.5	6.2	436	3	US-09-134-001C-3333
11	83.5	6.2	436	3	Sequence 2, Appl
12	83.5	6.2	436	3	Sequence 4, Appl
13	83.5	6.2	436	3	Sequence 6, Appl
14	83.5	6.2	436	3	Sequence 8, Appl
15	83	6.2	679	1	US-08-669-378-10
16	82.5	6.2	1162	4	US-08-441-139-5
17	82.5	6.2	2199	5	US-09-134-001C-4008
18	81	6.0	447	4	PCT-US95-11684-2
19	81	6.0	460	3	Sequence 2, Appl
20	81	6.0	460	3	Sequence 4, Appl
21	81	6.0	460	4	US-09-935-263-4
22	81	6.0	509	4	US-09-594-185-4
23	80.5	6.0	674	4	US-09-134-078-18
24	80.5	6.0	330	4	US-09-160-119-2
25	80.5	6.0	436	3	US-09-134-001C-3582
26	80	6.0	463	2	Sequence 3582, Ap
27	80	6.0	432	2	Sequence 12, Appl
					Sequence 6, Appl
					Sequence 4028, Ap

Sequence 33, Appl
Sequence 33, Appl
Sequence 30, Appl
Sequence 4794, Ap
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 2, Appl
Sequence 4817, Ap
Sequence 4347, Ap
Sequence 2983, Ap
Sequence 2, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 6, Appl
Sequence 3741, Ap
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-032-372-12
; Sequence 12, Application US/09032372
; Patent No. 6008337
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Yue, Henry
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: CELL CYCLE RELATED PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032.372
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0478 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1292898
US-09-032-372-12


```

RESULT 4
US-08-926-842B-13
; Sequence 13, Application US/08926842B
; Patent No. 6030807
; GENERAL INFORMATION:
; APPLICANT: Sa-No. 6030807ueira, Isabel
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,842B
; FILING DATE: 10-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-089 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Bacillus subtilis
; FEATURE:
; OTHER INFORMATION: /product= "arab"
US-08-926-842B-13

Query Match 6.4%; Score 86; DB 3; Length 560;
Best Local Similarity 20.4%; Pred. No. 0.8;
Matches 59; Conservative 47; Mismatches 107; Indels 76; Gaps 12;

QY 4 VLDVGNTVAVLGVFESEGLRQHRMETDRKHTDEYGLMYKQLLEHGLSFEDYKGIIVS 63
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 96 ILPIDSSGQPLCMPEYEEEPHYVVKLWKHHAQKHADRLNQIAEEGEAFQRYGKIS 155
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 64 S--VPPIM-----FALERMCKEYFKIKPLVVGPGIKTGLNIKE----- 101
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 156 SEMMIPKVMQIAEAPHYEAADRIIEADWIYVQLGSLKRSNCTAGYKAMWSEKAGYP 215
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 102 -----NP--REVGADRIVNAVAGIHLGSPLIIVDFGTATTYCYINEEKHYMGVIT 151
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 216 SDDFFKLNPSMKTITKDLGSG--IHSVGEKA-----GSLT-----EKMAKLTGLL- 260
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 152 PGMISAEALYSRAALKPRIETKPS--VVGKNT-----VSAHQSGIL 193
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 261 PGTAIV-AVANVDAHVSPVAVGITEPGKMLMIMGTSTCHVLLGEEVHIVPGMCGYVDNGIL 319
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 194 YGVVGQVEG-----IVKRMKEAKQEPKVIATGGLAKLISESN 232
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 320 PGYAGYEGAGSCVGDHDFVFKIVCPYPAYQEEAKEKNIGVHELLSEKAN 368
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

```

; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439, 313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 537
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-313-537

Query Match 6.4%; Score 86; DB 4; Length 1228;
Best Local Similarity 26.7%; Pred. No. 2.7;
Matches 50; Conservative 28; Mismatches 67; Indels 42; Gaps 10;

QY 69 IMFALERMCEKYFKIKPLVVGPGIKTGLNLIKYNPREVGADR-----IVNAV 115
Db 493 ILFGKKYERERYEKV---IKACALKKDLQLLEDGDLTVIGDRGTTLSGGOKARVNLARAV 549
QY 116 ---AGIHLXGSPLIIVDFGTA-----TTVCYINEEKHYMGVITPGIMISAEALYSRAAK 167
550 YQADAIYLLDDPLSAVDAEVSRLHFELCICQILHEK-----IT--ILVTHQLQYLKAAS 601
QY 168 LPRIETKPSVVGNKNTVSA-MOSGILYGVGVGVEGIVKRMKEEAKQEPKVIATGGI LAKL 226
Db 602 --QILILKDGKMWQKGTYTEFLKSGIDFG-----SLLAKKDNESEQ-PPVPGTPTLNR 652
QY 227 ISEESNV 233
Db 653 TFSSESV 659

RESULT 7

US-09-605-785-538
; Sequence 538, Application US/09605785
; Patent No. 6321716

GENERAL INFORMATION:

; APPLICANT: Xu, JIANGCHUN
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

DIAGNOSIS OF PROSTATE CANCER

FILE REFERENCE: 210121.427C16

CURRENT APPLICATION NUMBER: US/09/605,785

CURRENT FILING DATE: 2000-06-27

NUMBER OF SEQ ID NOS: 835

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 538

LENGTH: 1261

TYPE: PRT

ORGANISM: Homo sapiens

US-09-605-785-538

Query Match 6.4%; Score 86; DB 4; Length 1261;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 50; Conservative 28; Mismatches 67; Indels 42; Gaps 10;

QY 69 IMFALERMCEKYFKIKPLVVGPGIKTGLNLIKYNPREVGADR-----IVNAV 115

Db 450 ILFGKKYERERYEKV---IKACALKKDLQLLEDGDLTVIGDRGTTLSGGOKARVNLARAV 506
QY 116 ---AGIHLXGSPLIIVDFGTA-----TTVCYINEEKHYMGVITPGIMISAEALYSRAAK 167
Db 507 YQADAIYLLDDPLSAVDAEVSRLHFELCICQILHEK-----IT--ILVTHQLQYLKAAS 558
QY 168 LPRIETKPSVVGNKNTVSA-MOSGILYGVGVGVEGIVKRMKEEAKQEPKVIATGGI LAKL 226
Db 559 --QILILKDGKMWQKGTYTEFLKSGIDFG-----SLLAKKDNESEQ-PPVPGTPTLNR 609
QY 227 ISEESNV 233
Db 610 TFSSESV 616

RESULT 8

US-09-439-313-538
; Sequence 538, Application US/09439313
; Patent No. 6329505

GENERAL INFORMATION:

; APPLICANT: Xu, JIANGCHUN
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

DIAGNOSIS OF PROSTATE CANCER

FILE REFERENCE: 210121.427C9

CURRENT APPLICATION NUMBER: US/09/439, 313

CURRENT FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 575

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 538

LENGTH: 1261

TYPE: PRT

ORGANISM: Homo sapiens

US-09-439-313-538

Query Match 6.4%; Score 86; DB 4; Length 1261;

Best Local Similarity 26.7%; Pred. No. 2.8;

Matches 50; Conservative 28; Mismatches 67; Indels 42; Gaps 10;

QY 69 IMFALERMCEKYFKIKPLVVGPGIKTGLNLIKYNPREVGADR-----IVNAV 115

Db 450 ILFGKKYERERYEKV---IKACALKKDLQLLEDGDLTVIGDRGTTLSGGOKARVNLARAV 506

QY 116 ---AGIHLXGSPLIIVDFGTA-----TTVCYINEEKHYMGVITPGIMISAEALYSRAAK 167

Db 507 YQADAIYLLDDPLSAVDAEVSRLHFELCICQILHEK-----IT--ILVTHQLQYLKAAS 558

QY 168 LPRIETKPSVVGNKNTVSA-MOSGILYGVGVGVEGIVKRMKEEAKQEPKVIATGGI LAKL 226

Db 559 --QILILKDGKMWQKGTYTEFLKSGIDFG-----SLLAKKDNESEQ-PPVPGTPTLNR 609

QY 227 ISEESNV 233

Db 610 TFSSESV 616

RESULT 9

US-09-134-001C-3333

; Sequence 3333, Application US/09134001C

; Patent No. 6380370

GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

REGISTRATION NUMBER: 29,768

ATTORNEY/AGENT INFORMATION

APPLICATION NUMBER: DE P 44 00 926.7

NAME: DiGiglio, Frank S.

```
;
;  REGISTRATION NUMBER: 31,346
;  REFERENCE/DOCKET NUMBER: 8646
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: 516-742-4343
;  TELEFAX: 516-742-4366
;  TELEX: 230 901 SANS UR
;  INFORMATION FOR SEQ ID NO: 5:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 679 amino acids
;  TYPE: amino acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
;  US-08-441-139-5

Query Match      6.2%; Score 83; DB 1; Length 679;
Best Local Similarity 21.4%; Pred. No. 2.3;
Matches 55; Conservative 34; Mismatches 72; Indels 96; Gaps 13;

QY  43 VKOLLEHEGLSFEDVRGIIV---SSVPPIMFALERMCCKYFKIKPLVVGPGIKTGILNIK 99
    |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
363 VERVLKDSGLQKEDIDIVLVGGSTRIPKV---QQLLEKFFNGKK--ASKGINPDEAVA 416

QY  100 YENPREVGADRIYNAGIHLYGSPLIYD-----FGTATTCYIINEEKHYMGVITP-- 152
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  417 YGAAVQAG---VLSGEEGV---EDIVLLDVNALTGLIETT-----GGVMTPLI 458

QY  153 ----GIMISAEALYS-----RAAK--LPRI 171
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  459 KRNTAIPTKSQIFSTAVDNQKAVRIQVVEGERAMVKNLNLGNFELSDIRAAPRGVPQI 518

QY  172 EITKPSVVGKNTVSAMQSGILYGYGVQVEGI-----VCRMKEEAKQEPKVI 218
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  519 EVTFALDANGILIVSATDKD-----TGKSESITIANDKGLSQDDIDRWVEEAE---KYA 570

QY  219 ATGGLAKLISEESNVID 235
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  571 AEDAKFKAKSEARNTFE 587
```

Search completed: June 24, 2003, 22:10:40
Job time : 12.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.5605 Seconds
(without alignments)
1947.059 Million cell updates/sec

Title: US-09-813-453A-45
Perfect score: 1341
Sequence: 1 MIFVLDVGNNTNAVLGVFEEG.....LKGLYMLYERNALQHEKGE 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues 417779
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1341	100.0	262	9	US-09-813-453A-45
2	1057.5	78.9	258	9	US-09-813-453A-49
3	1034	77.1	258	9	US-09-813-453A-2
4	960	71.6	254	9	US-09-813-453A-47
5	901	67.2	233	9	US-09-813-453A-17
6	802.5	59.8	256	9	US-09-813-453A-55
7	783	58.4	255	9	US-09-813-453A-7
8	665.5	49.6	250	9	US-09-813-453A-3
9	638	47.6	265	9	US-09-813-453A-4
10	566	42.2	260	9	US-09-813-453A-51
11	536	40.0	258	9	US-09-813-453A-6
12	527	39.3	219	9	US-09-813-453A-57
13	471.5	35.2	272	9	US-09-813-453A-5
14	467.5	34.9	272	9	US-09-712-363-276
15	405.5	30.2	262	9	US-09-813-453A-8
16	399.5	29.8	246	9	US-09-813-453A-9
17	379.5	28.3	212	9	US-09-813-453A-59
18	342.5	25.5	273	9	US-09-813-453A-10
19	316	23.6	257	9	US-09-813-453A-53

20	276.5	20.6	262	9	US-09-813-453A-11	Sequence 11, Appl
21	207.5	15.5	244	9	US-09-813-453A-41	Sequence 41, Appl
22	190.5	14.2	241	9	US-09-813-453A-63	Sequence 63, Appl
23	169.5	12.6	592	9	US-09-813-453A-43	Sequence 43, Appl
24	168	12.5	460	9	US-09-813-453A-39	Sequence 39, Appl
25	168	12.5	592	9	US-09-813-453A-22	Sequence 22, Appl
26	167.5	12.5	229	9	US-09-813-453A-12	Sequence 12, Appl
27	156.5	11.7	257	9	US-09-813-453A-13	Sequence 13, Appl
28	142	10.6	249	9	US-09-813-453A-61	Sequence 61, Appl
29	138	10.3	267	9	US-09-813-453A-15	Sequence 15, Appl
30	137	10.2	249	9	US-09-813-453A-70	Sequence 70, Appl
31	136	10.1	248	9	US-09-813-453A-20	Sequence 20, Appl
32	117	8.7	242	9	US-09-813-453A-65	Sequence 65, Appl
33	113	8.4	209	9	US-09-813-453A-21	Sequence 21, Appl
34	98.5	7.3	223	9	US-09-895-913A-74	Sequence 74, Appl
35	98.5	7.3	223	9	US-09-813-453A-14	Sequence 14, Appl
36	98.5	7.3	223	9	US-09-813-453A-67	Sequence 67, Appl
37	97	7.2	509	10	US-09-815-242-5796	Sequence 5796, Ap
38	97	7.2	517	10	US-09-815-242-12837	Sequence 12837, A
39	90	6.7	294	10	US-09-815-242-13475	Sequence 13475, A
40	88.5	6.6	1019	1	US-08-834-705-18	Sequence 18, Appl
41	86	6.4	357	9	US-10-012-896-826	Sequence 826, App
42	86	6.4	357	9	US-09-895-793-826	Sequence 826, App
43	86	6.4	357	9	US-09-895-814-826	Sequence 826, App
44	86	6.4	357	10	US-09-759-143-826	Sequence 826, App
45	86	6.4	357	10	US-09-780-669-826	Sequence 826, App

ALIGNMENTS

RESULT 1
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match	100.0%	Score 1341;	DB 9;	Length 262;
Best Local Similarity	100.0%	Pred. No. 3.9e-121;		
Matches 262;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MIFVLDVGNNTNAVLGVFEEGELRQHWMETDRHKTEDEYGMVLMVKKLLEHGLSFDVKG	60	
Db	1	MIFVLDVGNNTNAVLGVFEEGELRQHWMETDRHKTEDEYGMVLMVKKLLEHGLSFDVKG	60	
QY	61	IVSSVVPIMPALERCEKYPKIKPLVVGPGTKTGLNLYENPREVGADRIYNAGVH	120	
Db	61	IVSSVVPIMPALERCEKYPKIKPLVVGPGTKTGLNLYENPREVGADRIYNAGVH	120	
QY	121	YGSPLIIVDFGTATTTCYINNEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSV	180	
Db	121	YGSPLIIVDFGTATTTCYINNEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSV	180	
QY	181	GKNTYSAMSGILYGVQVEGIVKRMKEAKQEPKVIATGGLAKLISESNVIDVDPF	240	
Db	181	GKNTYSAMSGILYGVQVEGIVKRMKEAKQEPKVIATGGLAKLISESNVIDVDPF	240	

Db 181 GNTVSAMQSGILYGVQGVGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIVDVPDF 240
QY 241 LTLKGLYMLYERNANLQHEKG 262
Db 241 LTLKGLYMLYERNANLQHEKG 262

RESULT 2

US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

Query Match 78.9%; Score 1057.5; DB 9; Length 258;
Best Local Similarity 77.0%; Pred. No. 7.4e-94;
Matches 201; Conservative 26; Mismatches 31; Indels 3; Gaps 1;
QY 1 MIFVLDVGNNTAVLGVEEGELRQHWMETDRHKTDEYGMVYKQLEHEGLSFDVKGI 60
Db 1 MIFVLDVGNNTAVLGVEEGELRQHWMETDRHKTDEYGMVYKQLEHEGLSFDVKGI 60
QY 61 IVSSVPPIMFALERCCEYFKIKPLVGVGPKTGLNINIKYENPREVGADRIYNVAGIHL 120
Db 61 IVSSVPPIMFALERCCEYFKIKPLVGVGPKTGLNINIKYENPREVGADRIYNVAGIHL 120
QY 121 YGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 121 YGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
QY 181 GNTVSAMQSGILYGVQGVGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIVDVPDF 240
Db 181 GNTVSAMQSGILYGVQGVGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIVDVPDF 240
241 LTLKGLYMLYERNANLQHEKG 261
241 LTLKGLYMLYERNANLQHEKG 262

RESULT 3

US-09-813-453A-49
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 77.1%; Score 1034; DB 9; Length 258;
Best Local Similarity 75.1%; Pred. No. 1.3e-91;
Matches 190; Conservative 34; Mismatches 29; Indels 0; Gaps 0;
QY 1 MIFVLDVGNNTAVLGVEEGELRQHWMETDRHKTDEYGMVYKQLEHEGLSFDVKGI 60
Db 1 MIFVLDVGNNTAVLGVEEGELRQHWMETDRHKTDEYGMVYKQLEHEGLSFDVKGI 60
QY 61 IVSSVPPIMFALERCCEYFKIKPLVGVGPKTGLNINIKYENPREVGADRIYNVAGIHL 120
Db 61 IVSSVPPIMFALERCCEYFKIKPLVGVGPKTGLNINIKYENPREVGADRIYNVAGIHL 120
QY 121 YGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 121 YGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
QY 181 GNTVSAMQSGILYGVQGVGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIVDVPDF 240
Db 181 GNTVSAMQSGILYGVQGVGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIVDVPDF 240
241 LTLKGLYMLYERN 253
241 LTLKGLYMLYERN 253

RESULT 4

US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 71.6%; Score 960; DB 9; Length 254;
Best Local Similarity 69.2%; Pred. No. 1.7e-84;
Matches 175; Conservative 35; Mismatches 43; Indels 0; Gaps 0;
QY 1 MIFVLDVGNNTAVLGVEEGELRQHWMETDRHKTDEYGMVYKQLEHEGLSFDVKGI 60
Db 1 MIFVLDVGNNTAVLGVEEGELRQHWMETDRHKTDEYGMVYKQLEHEGLSFDVKGI 60
QY 61 IVSSVPPIMFALERCCEYFKIKPLVGVGPKTGLNINIKYENPREVGADRIYNVAGIHL 120
Db 61 IVSSVPPIMFALERCCEYFKIKPLVGVGPKTGLNINIKYENPREVGADRIYNVAGIHL 120
QY 121 YGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 121 YGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
QY 181 GNTVSAMQSGILYGVQGVGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIVDVPDF 240
Db 181 GNTVSAMQSGILYGVQGVGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIVDVPDF 240

Db 181 GNTIDSMSGIFGYVSDVGVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240
QY 241 LTLKGLYLYERN 253
Db 241 LTLKGLYLYERN 253
RESULT 5
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-17

Query Match 67.2%; Score 901; DB 9; Length 233;
Best Local Similarity 75.5%; Pred. No. 7.4e-79;
Matches 163; Conservative 29; Mismatches 24; Indels 0; Gaps 0;
QY 1 MIFVLDVGNNTAVLGVEEGELRQHWRTEDRHKTEDEYGMVLVKOLLEHGLSFEEDVKGI 60
Db 1 MLLVFDVGNNTAVLGVEEGELRQHWRTEDRHKTEDEYGMVLVKOLLEHGLSFEEDVKGI 60
QY 61 IVSSVPPIMFALERCKEYFKIKPLVVGPGIKTGLNLYENPREVGADRIYNAVAGIHL 120
Db 61 IISVVPPIMFALERCKEYFKIEHPQIVGPGMKTGLNLYENPREVGADRIYNAVAGIHL 120
QY 121 YGSLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETPKSSVV 180
Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETPKSSVV 180
QY 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETPKSSVV 180
Db 121 YGNPLIVDFGTATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETPKSSVV 180
QY 181 GNTVSAMQSGILYGVQGVGIVKRMKEAKQEPK 216
Db 181 GNTVSAMQSGILYGVQGVGIVKRMKEAKQEPK 216

RESULT 6
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256

; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55
Query Match 59.8%; Score 802.5; DB 9; Length 256;
Best Local Similarity 59.1%; Pred. No. 2.6e-69;
Matches 150; Conservative 49; Mismatches 54; Indels 1; Gaps 1;
QY 1 MIFVLDVGNNTAVLGVEEGELRQHWRTEDRHKTEDEYGMVLVKOLLEHGLSFEEDVKGI 60
Db 1 MLLVFDVGNNTAVLGVEEGELRQHWRTEDRHKTEDEYGMVLVKOLLEHGLSFEEDVKGI 60
QY 61 IVSSVPPIMFALERCKEYFKIKPLVVGPGIKTGLNLYENPREVGADRIYNAVAGIHL 120
Db 61 IISVVPPIMFALERCKEYFKIKPLVVGPGIKTGLNLYENPREVGADRIYNAVAGIHL 120
QY 121 YGSLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETPKSSVV 180
Db 121 YGNPLIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETPKSSVV 180
QY 181 GNTVSAMQSGILYGVQGVGIVKRMKEAKQEPKVIATGGLAKLIGTESETIDVIDSF 239
Db 181 CKSTVSAMQSGILYGVQGVGIVKRMKEAKQEPKVIATGGLAKLIGTESETIDVIDSF 240
QY 240 FTLKGLYLYERN 253
Db 241 FTLKGLYLYERN 254

RESULT 7
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 58.4%; Score 783; DB 9; Length 255;
Best Local Similarity 57.3%; Pred. No. 1.9e-67;
Matches 145; Conservative 51; Mismatches 57; Indels 0; Gaps 0;
QY 1 MIFVLDVGNNTAVLGVEEGELRQHWRTEDRHKTEDEYGMVLVKOLLEHGLSFEEDVKGI 60
Db 1 MLLVFDVGNNTAVLGVEEGELRQHWRTEDRHKTEDEYGMVLVKOLLEHGLSFEEDVKGI 60
QY 61 IVSSVPPIMFALERCKEYFKIKPLVVGPGIKTGLNLYENPREVGADRIYNAVAGIHL 120
Db 61 IISVVPPIMFALERCKEYFKIKPLVVGPGIKTGLNLYENPREVGADRIYNAVAGIHL 120
QY 121 YGSLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETPKSSVV 180
Db 121 YGSLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETPKSSVV 180
QY 181 GNTVSAMQSGILYGVQGVGIVKRMKEAKQEPKVIATGGLAKLIGTESETIDVIDSF 240
Db 181 ARNTVSMQAGIYGVGLVDEIVTRMKAESDAPRVATGGLASLIAPESKTEAVEST 240

```

US-09-813-453A-4
Sequence 4, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Paterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS
TITLE OF INVENTION: ANTIBIOTICS
FILE REFERENCE: OGG-001
CURRENT APPLICATION NUMBER: US/09/813-453A-4
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,565
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 265
TYPE: PRT
ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

```

Qy	1	MIFVLVDGNTNAVGLVFEEGELRQHWRMETDRHKTEDEYGLMVKQLLEHGEGLSTEDYKGI	60
Db	1	MLLAIEQGNNTFMFAIHGDGASVVAQWRSATETSTRTADEYVYVWLSQLLSMQGLSPRAIDAY	60
Qy	61	IVSSVVPPTMFALERCCKYFKIKPLVYVPGTKTGLNIKYENPREVGADRTVNAVAGTGL	120
Db	61	IISVVVQSFIMRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPEAGADRLVNAIGAAMV	120
Qy	121	YGSPILIVDFGPTATTCYINEKKHYMGVITPGIMISAEALYSRAAKLPRTEITKPS--	170
Db	121	YPGPLVVIDSGPTATTFDVIADGAPGEIGTIIAPGNLSMQALHEAAAKLPRIAIQRPAGNR	180
Qy	179	VYGGKNTVSAQSGILYGVVGVEGIVTKRMKEAKOEKVIATGTGLAKLISESNVIDVD	230
Db	181	IVGDTVSAQSGVFWGIVISLIEGLVARIKAEKPEPTVIATGGVSLFEGATSIDHFD	240
Qy	239	PFLTKLKGILMYERN	253
Db	241	SDLITIRGLLEIYRRN	255


```

; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 6027,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Deinococcus radiopugnans
US-09-813-453A-8

Query Match 30.2%; Score 405.5
Best Local Similarity 31.7%; Pred. No. 4
Matches 84; Conservative 60; Mismatch

QY 2 IFVLVGNVTNAVLCVFE-BGELRQHKWRMETDID
Db : : : : : : : : : : : : : : : :
6 LLAVDIGNTTVLGLADASGALHTHTWRIRTN
QY 61 IVSSVPPPTMFALERCKEYFKIKPLVPGV
Db : : : : : : : : : : : : : : : :
64 VLSSVAPPGVENYALALKRHFMDAFVSAEAE
QY 121 YGSP-----LIIVDFGTATTTCYINEK
Db : : : : : : : : : : : : : : : :
118 FGAEKYLGGLDYAVVYDFGTSTNFVDVGRGR
QY 173 IYKPSVVGKNTVSAMOSGILYGVQVQSGI
Db : : : : : : : : : : : : : : : :
178 LQAPETAIGKNTVHALQSLGVFGYAEWVDGL
QY 233 VLDVDPFPTLKGLYMLYERNANLQ 257
Db : : : : : : : : : : : : : : : :
238 EIDYDETLLRGLVELWASRSEVR 262

Search completed: June 24, 2003, 22:29:00
Job time : 15.5605 secs

```

Search completed: June 24, 2003, 22:29:00
Job time : 15.5605 secs

```

RESULT 15
US-09-813-453A-8
; Sequence 8, Application US/09813453A
; Patent No. US20020168681a1
; GENERAL INFORMATION
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: CGZ-001

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 29.657 Seconds

(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453A-45

Perfect score: 1341

Sequence: 1 MIFVLDVGNNTAVNLGVFEFG.....LKGLVLMYERNANLQHEKGE 262

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1341	100.0	262	AAU91170	Pantothenate kinas
2	1057.5	78.9	258	AAU91172	Pantothenate kinas
3	1034	77.1	258	AAU01243	B. subtilis novel
4	1034	77.1	258	AAU91149	Bacillus subtilis
5	960	71.6	254	AAU91171	Pantothenate kinas
6	901	67.2	233	AAU91163	Pantothenate kinas
7	872	65.0	259	ABB47661	Listeria monococog
8	802.5	59.8	256	AAU91175	Pantothenate kinas
9	783	58.4	255	AAU91154	Geobacter sulfurre
10	665.5	49.6	250	AAU91150	Clostridium acetob

11	638	47.6	265	23	AAU91151	Streptomyces coeli
12	566	42.2	260	23	AAU91173	Pantothenate kinas
13	536	40.0	258	23	AAU91153	Rhodobacter capsul
14	527	39.3	219	23	AAU91176	Pantothenate kinas
15	471.5	35.2	272	23	AAU91152	Mycobacterium tube
16	467.5	34.9	272	22	AAU91155	Deinococcus radiop
17	405.5	30.2	262	23	AAU91156	Thermotoga maritim
18	399.5	29.8	246	23	AAU91157	Pantothenate kinas
19	379.5	28.3	212	23	AAU91177	Treponema pallidum
20	342.5	25.5	273	23	AAU91157	Pantothenate kinas
21	316	23.6	257	23	AAU91174	Borrelia burgdorfe
22	276.5	20.6	262	23	AAU91158	Pantothenate kinas
23	207.5	15.5	244	23	AAU91168	Pantothenate kinas
24	190.5	14.2	241	23	AAU91179	Pantothenate kinas
25	169.5	12.6	389	21	AAU74909	Neisseria meningit
26	169.5	12.6	592	20	AAU38615	Neisseria meningit
27	169.5	12.6	592	21	AAU74912	Neisseria meningit
28	169.5	12.6	592	23	AAU91169	Pantothenate kinas
29	168	12.5	455	20	AAU38617	Neisseria gonorrh
30	168	12.5	455	21	AAU74908	Neisseria gonorrh
31	168	12.5	455	21	AAU74910	Neisseria meningit
32	168	12.5	460	23	AAU91167	Pantothenate kinas
33	168	12.5	592	20	AAU38618	Neisseria gonorrh
34	168	12.5	592	20	AAU38616	Neisseria meningit
35	168	12.5	592	21	AAU74911	Neisseria gonorrh
36	168	12.5	592	21	AAU74913	Neisseria meningit
37	168	12.5	592	23	AAU91166	Pantothenate kinas
38	167.5	12.5	229	23	AAU91159	Aquifex aeolicus p
39	156.5	11.7	257	23	AAU91178	Synechocystis pant
40	142	10.6	249	23	AAU91160	Pantothenate kinas
41	138	10.3	267	23	AAU91178	Bordella pertussis
42	137	10.2	249	23	AAU91182	Pantothenate kinas
43	136	10.1	248	23	AAU91164	Pantothenate kinas
44	118.5	8.8	189	20	AAU38614	Neisseria meningit
45	117	8.7	242	23	AAU91180	Pantothenate kinas

ALIGNMENTS

RESULT 1
AAU91170
ID AAU91170 standard; Protein; 262 AA.
XX
AC AAU91170;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #8.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
OS Bacillus anthracis.
XX
PN WO200216601-A2.
XX
PD 28-FEE-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
DR WPI: 2002-269358/31.
XX N-PSDB; ABR54191.
PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein
 PS Claim 8; Page 98-99; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 262 AA;
 Query Match 100.0%; Score 1341; DB 23; Length 262;
 Best Local Similarity 100.0%; Pred. No. 8.6e-135;
 Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 MIFVLDVGNNAVGLVFEFEGELRQHWRMETDRHKTDEYGMVQLLEHGLSFEDVKGI 60
 1 MIFVLDVGNNAVGLVFEFEGELRQHWRMETDRHKTDEYGMVQLLEHGLSFEDVKGI 60
 QY 61 IVSSVPPIMFALERMCEKYFKIKPLVGVGPKTGLNRYENPREVGADRIYNVAVAGIHL 120
 DB 61 IVSSVPPIMFALERMCEKYFKIKPLVGVGPKTGLNRYENPREVGADRIYNVAVAGIHL 120
 QY 121 YGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
 DB 121 YGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
 QY 181 GNTVSAMQSGILYGVGVGEGIVKRMKEAKOEKPKVIATGGGLAKLISESNVIDVDDPF 240
 DB 181 GNTVSAMQSGILYGVGVGEGIVKRMKEAKOEKPKVIATGGGLAKLISESNVIDVDDPF 240
 QY 241 LTLKGLYMLYERNANLQHEKG 262
 DB 241 LTLKGLYMLYERNANLQHEKG 262
 RESULT 2
 AAU91172
 ID AAU91172 standard; Protein; 258 AA.
 XX
 AC AAU91172;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 PT Pantothenate kinase (Coax) #10.
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX
 OS Bacillus stearothermophilus.
 XX
 PN WO200216601-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26531.
 XX
 PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.
 DR N-PSDB; ABK54193.
 XX

PT Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein
 XX
 PS Claim 10; Page 101-102; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 258 AA;
 Query Match 78.9%; Score 1057.5; DB 23; Length 258;
 Best Local Similarity 77.0%; Pred. No. 1.7e-104;
 Matches 201; Conservative 26; Mismatches 31; Indels 3; Gaps 1;
 QY 1 MIFVLDVGNNAVGLVFEFEGELRQHWRMETDRHKTDEYGMVQLLEHGLSFEDVKGI 60
 DB 1 MIFVLDVGNNAVGLVFEFEGELRQHWRMETDRHKTDEYGMVQLLEHGLSFEDVKGI 60
 QY 61 IVSSVPPIMFALERMCEKYFKIKPLVGVGPKTGLNRYENPREVGADRIYNVAVAGIHL 120
 DB 61 IVSSVPPIMFALERMCEKYFKIKPLVGVGPKTGLNRYENPREVGADRIYNVAVAGIHL 120
 QY 121 YGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
 DB 121 YGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
 QY 181 GNTVSAMQSGILYGVGVGEGIVKRMKEAKOEKPKVIATGGGLAKLISESNVIDVDDPF 240
 DB 181 GNTVSAMQSGILYGVGVGEGIVKRMKEAKOEKPKVIATGGGLAKLISESNVIDVDDPF 240
 QY 241 LTLKGLYMLYERNANLQHEKG 261
 DB 241 LTLKGLYMLYERNANLQHEKG 258
 RESULT 3
 AAU01243
 ID AAU01243 standard; Protein; 258 AA.
 XX
 AC AAU01243;
 XX
 DT 18-JUL-2001 (first entry)
 XX
 PT B. subtilis novel pantothenate kinase encoded by the gene coax.
 DE Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;
 KW nutritional supplement; panto-compound; pantoate.
 KW
 XX Bacillus subtilis.
 OS
 PN WO200121772-A2.
 XX
 PD 29-MAR-2001.
 XX
 PF 21-SEP-2000; 2000WO-US25993.
 XX
 PR 21-SEP-1999; 99US-0400494.
 PR 07-JUN-2000; 2000US-0210072.
 PR 28-JUL-2000; 2000US-0221836.
 PR 24-AUG-2000; 2000US-0227860.
 XX
 XX (OMNI-) OMNIGENE BIOPRODUCTS.
 PA Yocum RR, Patterson TA, Hermann T, Pero JG;
 PI

PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54192.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
PS Claim 10; Page 100; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 254 AA;
Query Match 71.6%; Score 960; DB 23; Length 254;
Best Local Similarity 69.2%; Pred. No. 4.5e-94;
Matches 175; Conservative 35; Mismatches 43; Indels 0; Gaps 0;
QY 1 MIFVLVDGNTNAVGLVFEEGELRQHWMETDRHKTDEYGMVLKQLEHGLSFEDVKGI 60
DB 1 MLIVLDVGNNTVLGVYQDETLLVHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQIDGI 60
QY 61 IVSSVPPIMFALERMCEYFKIKPLVVGPGIKTGLNINIKYENPREVGADRIYNVAGIHL 120
DB 61 VISSVPPMFMSELEQCKKYFHTVPMIIGIKTGLNINIKYDNPKYEGADRIYNVAAIEL 120
QY 121 YGSPILIVDFGATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
DB 121 YGPAIVVDGATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
QY 181 GNTVSAMQSGLYGVGVGEGIVKRMKEEAKOEKPKVATGGLAKLISESNVIDVDF 240
DB 181 GNTIDMSOGIFGYVSQDGVVKKRKAQAESEPKVIATGGLAKLIGTESETIDVDSF 240
241 LTLKGLYLYERN 253
241 LTLKGLQLYKKN 253
RESULT 6
AAU91163
ID AAU91163 standard; Protein; 233 AA.
XX
AC AAU91163;
XX
DT 05-JUN-2002 (first entry)
DE
DE Pantothenate kinase (Coax) #1.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus subtilis.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX

PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54169.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
PS Disclosure; Page 81-82; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 233 AA;
Query Match 67.2%; Score 901; DB 23; Length 233;
Best Local Similarity 75.5%; Pred. No. 8e-88;
Matches 163; Conservative 29; Mismatches 24; Indels 0; Gaps 0;
QY 1 MIFVLVDGNTNAVGLVFEEGELRQHWMETDRHKTDEYGMVLKQLEHGLSFEDVKGI 60
DB 1 MLIVLDVGNNTVLGVYHDKLEYHWRIETSRHKTDEFGMLRSFLDHSGLMFEQIDGI 60
QY 61 IVSSVPPIMFALERMCEYFKIKPLVVGPGIKTGLNINIKYENPREVGADRIYNVAGIHL 120
DB 61 VISSVPPIMFALERMCEYFKIKPLVVGPGIKTGLNINIKYDNPKYEGADRIYNVAAIHL 120
QY 121 YGSPILIVDFGATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
DB 121 YGNPLIVDFGATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
QY 181 GNTVSAMQSGLYGVGVGEGIVKRMKEEAKOEKPK 216
DB 181 GNTVSAMQSGLYGVGVGEGIVKRMKEEAKOEKPK 216
RESULT 7
ABB47661
ID ABB47661 standard; Protein; 259 AA.
XX
AC ABB47661;
XX
DT 05-FEB-2002 (first entry)
DE
DE Listeria monocytogenes protein #365.
XX
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
XX WO200177335-A2.
XX
XX 18-OCT-2001.
XX
XX 11-APR-2001; 2001WO-FR01118.
XX

PR 11-APR-2000; 2000FR-0004629.
XX (INSP) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Deboux P;
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
PT and prevention of *Listeria* and related bacterial infections, and
PT related polypeptides -
XX Claim 6; SEQ ID No 366; 192pp; French.
XX The present invention relates to the genome sequence of *Listeria*
CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in *L.*
CC *monocytogenes* and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of *L. monocytogenes* and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by *L.*
CC *monocytogenes* and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 259 AA;
Query Match 65.0%; Score 872; DB 23; Length 259;
Best Local Similarity 61.7%; Pred. NO. 1.2e-84;
Matches 156; Conservative 49; Mismatches 48; Indels 0; Gaps 0;
QY 1 MIFVLDVGNNAVGLGVFEELRQHRMETDRHKTDEYGLMVKQLLEHGLSFDVKG 60
Db 1 MLVLDVGNNTCTVGYEKQLLKHWRMTDRHRTSDELGMTVLNFFSYANLTPSDIQI 60
QY 61 IVSSVVPPIPMFALERCKEYFKIKPLVVGPGIKTGLNLYENPREVGADRIYNAVAGIHL 120
Db 61 IISVVPPIPMAMETVCVRYNIRPLVVGPGIKTGLNLYENPREIGSDRIYNAVAASEE 120
QY 121 YGSPLIIVDFGTATTCYCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 121 YGTPVIVDFGTATTCYCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIYDIAESSQII 180
QY 181 GKNTVSAMQSGILYGVGOVGVIGVYKRMKEAKQEPKVTATGGLAKLISEESNVIDVDPF 240
Db 181 GKSTVSSMQAGIFYGVGOCEGIIAEMKKQSNAPVVVATGGLARMITEKSAVDILDPF 240
QY 241 LTLKGLMYLYERN 253
Db 241 LTLKGLLYLYERN 253
RESULT 8
AAU91175
ID AAU91175 standard; Protein; 256 AA.
XX
AC AAU91175;
XX

DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #13.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Clostridium difficile.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABR54196.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
PS Claim 6; Page 105; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 256 AA;
Query Match 59.8%; Score 802.5; DB 23; Length 256;
Best Local Similarity 59.1%; Pred. NO. 3.1e-77;
Matches 150; Conservative 49; Mismatches 54; Indels 1; Gaps 1;
QY 1 MIFVLDVGNNAVGLGVFEELRQHRMETDRHKTDEYGLMVKQLLEHGLSFDVKG 60
Db 1 MLVLDVGNNTCTVGYEKQLLKHWRMTDRHRTSDELGMTVLNFFSYANLTPSDIQI 60
QY 61 IVSSVVPPIPMFALERCKEYFKIKPLVVGPGIKTGLNLYENPREVGADRIYNAVAGIHL 120
Db 61 IISVVPPIPMAMETVCVRYNIRPLVVGPGIKTGLNLYENPREIGSDRIYNAVAASEE 120
QY 121 YGSPLIIVDFGTATTCYCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 121 YGTPVIVDFGTATTCYCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIYDIAESSQII 180
QY 181 GKNTVSAMQSGILYGVGOVGVIGVYKRMKEAKQEPKVTATGGLAKLISEESNVIDVDPF 239
Db 181 GKSTVSSMQAGIFYGVGOCEGIIAEMKKQSNAPVVVATGGLARMITEKSAVDILDPF 240
QY 241 LTLKGLMYLYERN 253
Db 241 LTLKGLLYLYERN 253
RESULT 9
AAU91154
ID AAU91154 standard; Protein; 255 AA.
XX

AAU91154;

05-JUN-2002 (first entry)

Geobacter sulfurreducens pantothenate kinase Coax.

Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.

Geobacter sulfurreducens.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

Identifying potential antibiotic or antimicrobial agent, comprises
contacting composition comprising pantothenate kinase (Coax) protein
with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 72-73; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic
comprising contacting an assay composition comprising a pantothenate
kinase (Coax) protein with a test compound, and determining the ability
of the test compound to inhibit the activity of the Coax protein, an
essential enzyme for the production of coenzyme A. Coax protein is a
valuable target for identifying bactericidal compounds. Coax modulating
agents can be used in an infectious animal model to determine the
efficacy, toxicity, or side effects of treatment with such an agent. This
is the amino acid sequence of a pantothenate kinase (Coax) protein
described in the invention.

Sequence 255 AA;

Query Match 58.4%; Score 783; DB 23; Length 255;

Best Local Similarity 57.3%; Pred. No. 3.8e-75;

Matches 145; Conservative 51; Mismatches 57; Indels 0; Gaps 0;

1 MIFVLDVGNNTNAVIGVFEEGELRQHRMETDRHKTDEYGMVQLLEHGLSFEDVKG 60

1 MLLVIDVGNNTNVLGIYDNGRLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60

61 IVSSVPPIMFALERCCEKFKIKPLVVGIGIKTGLNIKYENPREVGADRIVNAVAGIHL 120

61 IISVVPPLGVLSLGYFGRPLVVGIGIKTGMPIQYDNPREGVADRIVNAVAGYEK 120

121 YGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETTKPSSV 180

121 YRTSLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETTKPSSV 180

181 GKNVTSAMQSGILYGVGVGVEGIVKRMKEAKPEKVIATGGGLAKLISESNVIDVDPF 240

181 ARNTVNSMQAGIYGVGVGLVDEIVTRMAESKDPAPRVATGGGLASLIAPESKTIIEVEEY 240

241 LTLKGLMYLERN 253

241 LTLEGLRILYERN 253

RESULT 10

AAU91150

ID AAU91150 standard; Protein; 250 AA.

XX

AC

XX

DT

XX

DE

XX

KW

XX

KW

XX

OS

XX

PN

XX

PD

XX

XX

PF

XX

PR

XX

PR

XX

PA

XX

PI

XX

DR

XX

XX

XX

PT

XX

PT

XX

PT

XX

PS

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

AAU91150;

05-JUN-2002 (first entry)

Clostridium acetobutylicum pantothenate kinase Coax.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Clostridium acetobutylicum.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

24-AUG-2000; 2000US-227860P.

20-MAR-2001; 2001US-0813453.

(OMNI-) OMNIGENE BIOPRODUCTS INC.

Yocum RR, Patterson TA;

WPI; 2002-269358/31.

Identifying potential antibiotic or antimicrobial agent, comprises
contacting composition comprising pantothenate kinase (Coax) protein
with test compound and identifying inhibitor of the Coax protein -

Claim 10; Page 68-69; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic
comprising contacting an assay composition comprising a pantothenate
kinase (Coax) protein with a test compound, and determining the ability
of the test compound to inhibit the activity of the Coax protein, an
essential enzyme for the production of coenzyme A. Coax protein is a
valuable target for identifying bactericidal compounds. Coax modulating
agents can be used in an infectious animal model to determine the
efficacy, toxicity, or side effects of treatment with such an agent. This
is the amino acid sequence of a pantothenate kinase (Coax) protein
described in the invention.

Sequence 250 AA;

Query Match 49.6%; Score 665.5; DB 23; Length 250;

Best Local Similarity 54.9%; Pred. No. 1.3e-82;

Matches 128; Conservative 45; Mismatches 55; Indels 5; Gaps 1;

1 MIFVLDVGNNTNAVIGVFEEGELRQHRMETDRHKTDEYGMVQLLEHGLSFEDVKG 60

18 VILVLDVGNNTNVLGIYDNTKLTAEWRLSTDVLRSADEYGIQVWNLFOQDKLPTLVEGV 77

61 IVSSVPPIMFALERCCEKFKIKPLVVGIGIKTGLNIKYENPREVGADRIVNAVAGIHL 120

78 IISVVPNTMYSLEHMRKFKINPLVVGIGIKTGLNIKYENPREVGADRIVNAVAAHEI 137

121 YGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETTKPSSV 180

138 YKRSLLIIDFGTATTCFAVRENGDYLGGACPCGIVSSSEALFEKAKLPVELIKPAYAI 197

181 GKNVTSAMQSGILYGVGVGVEGIVKRMKEAKPEK-----VTATGLAKLIS 228

198 CKNTISSIOSGIVRYLRQVKLFKELKENLPDGRTRTSLVATGLAKLIN 250

RESULT 11

AAU91151

ID AAU91151 standard; Protein; 265 AA.

DT 05-JUN-2002 (first entry).
DE Streptomyces coelicolor pantothenate kinase Coax.
DE Streptomyces coelicolor pantothenate kinase Coax.
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Streptomyces coelicolor.
OS WO200216601-A2.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PI Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX Claim 10; Page 69-70; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX Sequence 265 AA;
SQ
Query Match 47.6%; Score 638; DB 23; Length 265;
Best Local Similarity 47.9%; Pred. No. 1.3e-59;
Matches 126; Conservative 55; Mismatches 74; Indels 8; Gaps 3;
1 MIFVLDVGNNTNAVGLVFEFEGELRQHRMETDRHKTDEYGLVQLLEHGLSPED---- 56
1 MLLTIDVGNTHVLGDFDGEDIVEHWRISTDSRRTADELAVLLQGLMGHPHLLGDELGDG 60
57 VKGIIVSSVVPPIAFALERMCEKYFKIKPLV-VGPGIKTGLNRYENPREVGADRIYNAV 115
61 IDGTAICATVPSVLHELREVTTRYIGDVPALVPEGVATGVPILTDHPKEVGADRIYNAV 120
116 AGIHLVGLPIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETTK 175
121 AAELYGGPAIVDFGTATTEDAVSARGEYIGGVATGIEISVEALGVKGALRKIEVAR 180
176 PSSVVGKNTVSAMQSGIYGVGVQVEGIVKRMKEAKQEP---KVIATGGGLAKLISEB 232
181 PRSIVGKNTVSAMQSGIYGVGVQVEGIVKRMKEAKQEP---KVIATGGGLAKLISEB 232
233 VIDVVDPLTLKGLYMLYERNAN 255
241 VIDEHEPWITLMLGLRYERNVS 263
RESULT 12:
AAU91173
ID AAU91173 standard; Protein; 260 AA.
XX
AC AAU91173;

XX 05-JUN-2002 (first entry)
DT Pantothenate kinase (Coax) #11.
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Caulobacter crescentus.
OS WO200216601-A2.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
PI Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX N-PSDB; ABK54194.
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX Claim 10; Page 102-103; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX Sequence 260 AA;
SQ
Query Match 42.2%; Score 566; DB 23; Length 260;
Best Local Similarity 42.0%; Pred. No. 6.2e-52;
Matches 107; Conservative 55; Mismatches 91; Indels 2; Gaps 1;
1 MIFVLDVGNNTNAVGLVFEFEGELRQHRMETDRHKTDEYGLVQLLEHGLSPEDVKG 60
1 MLLATEQGNNTMFAHDGASVVAQWRSATSTRTADEYVWVLSQLLSQGLGFRADAV 60
61 IVSSVVPPIAFALERMCEKYFKIKPLVVGPGIKTGLNRYENPREVGADRIYNAVAGIHL 120
61 IISVVPQSIIFNLRLNRRYFNVEPLVIGENAKLGDIVRIEKPSPSAGADRLVNAIGAAV 120
121 YGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETTKPSS-- 178
121 YPGPLVVIDSGTATTFDVAADGAGEGGLIAPGINSQALHEAAKLPRIATQIPAGNR 180
179 VVGKNTVSAMQSGIYGVGVQVEGIVKRMKEAKQEPKVIATGGGLAKLISEBNSVIDVD 238
181 IVGDTTVSAMQSGVFWGVIISLIEGLVARIKAEERGEPTVIATGGVASFEGATSDIDHD 240
239 PFLTLKGLYMLYERN 253
241 SDLATIRGLLEIYRRN 255
RESULT 13
AAU91153
ID AAU91153 standard; Protein; 258 AA.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.8087 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453A-45
Perfect score: 1341
Sequence: 1 MIFVLDVGNNTNAVIGVFEEG.....LKGLYMLYERNANLQHEKGE 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	960	71.6	254	2 F83660	hypothetical prote
2	901	67.2	233	2 S66100	conserved hypot
3	872	65.0	259	2 A11102	conserved hypot
4	867	64.7	259	2 A11464	conserved hypot
5	804	60.0	273	2 E97293	probable transcrip
6	638	47.6	265	2 T36391	hypothetical prote
7	566	42.2	261	2 B87489	transcription acti
8	492.5	36.7	274	2 H86937	conserved hypot
9	471.5	35.2	272	2 A70955	hypothetical prote
10	405.5	30.2	262	2 E75516	conserved hypot
11	399.5	29.8	246	2 D72320	conserved hypot
12	342.5	25.5	273	2 D71326	conserved hypot
13	276.5	20.6	262	2 F70165	conserved hypot
14	169.5	12.6	592	2 B81009	Bira protein/Bvg a
15	168	12.5	592	2 H82031	probable biotin-la
16	167.5	12.5	229	2 E70465	hypothetical prote
17	156.5	11.7	257	2 S75559	hypothetical prote
18	147	11.0	276	2 A12292	hypothetical prote
19	138	10.3	267	2 I40327	baf protein - Bord
20	136	10.1	248	2 H83111	hypothetical prote
21	121	9.0	224	2 A95571	conserved hypot
22	117	8.7	242	2 A82637	conserved hypot
23	113	8.4	209	2 H81382	hypothetical prote
24	112.5	8.4	223	2 G71887	hypothetical prote
25	112.5	8.4	324	2 F90311	hypothetical prote
26	103.5	7.7	351	2 F85357	GDP-mannose pyroph
27	103.5	7.7	597	2 B89251	probable electron
28	103.5	7.7	736	2 G72621	probable translati
29	98.5	7.3	223	2 F64627	hypothetical prote

30	97.5	7.3	400	2 H82833	proline dipeptidas
31	96.5	7.2	828	2 G82583	bifunctional aspar
32	96	7.2	933	1 S56050	replication licens
33	95	7.1	1026	2 G90876	probable tail leng
34	93	6.9	219	2 E70348	UDP-3-O-[3-hydroxy
35	92.5	6.9	361	2 S67590	mannose-1-phosphat
36	92	6.9	317	2 S74030	hypothetical prote
37	92	6.9	592	2 G69252	aldehyde ferredoxi
38	92	6.9	1026	2 H85692	probable tail comp
39	92	6.9	1080	2 H90908	probable tail leng
40	91.5	6.8	1083	2 T18298	pyridine nucleotid
41	90.5	6.7	1084	2 T18292	nicotinamide nucle
42	90	6.7	294	2 A95195	ROK family protein
43	90	6.7	294	2 F98061	glucokinase [impor
44	90	6.7	323	2 F95390	probable threonine
45	90	6.7	428	2 A97341	adenylosuccinate s

ALIGNMENTS

RESULT 1

F83660
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83660
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: F83660
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:Cross-references: GB:AP001507; GB:BA000004; NID:gl0172612; PIDN:BA803805.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0086
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 71.6%; Score 960; DB 2; Length 254;
Best Local Similarity 69.2%; Pred. No. 2.6e-69;
Matches 175; Conservative 35; Mismatches 43; Indels 0; Gaps 0;

QY	1	MIFVLDVGNNTNAVIGVFEEGELRQHRMETDRHKTDEYGLVVKOLLEHEGLSFEDVKGI	60
DB	1	MILVIDVGNNTVLGVYQDETIVHHRLATSKQTEDEYAMTVRSFLDHAGLQFQDIDGI	60
QY	61	IVSSVVPPTMFALERCKEYFKIKPLVVGPGIKTGLNIKYENPREVGADRIYNAVAGIHL	120
DB	61	VISSVVPPTMFLEQMKCKYFHVTPMIGPGIKTGLNIKYDNPKVKGADRIYNAVAIEL	120
QY	121	YGSLPIIVDFGATTTCYINEEKHYMGVITFGIMISAEALYSRAAKLPRIETKPSVV	180
DB	121	YGPAIVVDFGATTTCYINEEKHYMGVITFGIMISAEALYSRAAKLPRIETKPSVV	180
QY	181	GNTYSAMSGILYGVGVEGIVKRMKEAKQEPKIATGGLAKLISSESNVIDVDPF	240
DB	181	GNTIDSMQSGIFGYVSQDGVVVRKKAQSEPKVIATGGLAKLIGTESETIDVDSF	240
QY	241	LTLKGLYMLYERN	253
DB	241	LTLKGLQLIYKKN	253

RESULT 2

S66100
conserved hypothetical protein yacB - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C:Accession: S66100; E69740
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.

H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H86937
R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A: Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A: Title: Massive gene decay in the leprosy bacillus.
A: Reference number: A86909; MUID: 21128732; PMID: 11234002
A: Accession: H86937
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-274 <STO>
A: Cross-references: GB:AL450380; NID: g13092576; PIDN: CAC29740.1; GSPDB: GN00147
A: Gene: ML0232
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 36.7%; Score 492.5; DB 2; Length 274;
Best Local Similarity 36.1%; Pred. No. 6.1e-32;
Matches 97; Conservative 67; Mismatches 90; Indels 15; Gaps 4;

QY 1 MIFVLDVGNNTNAVIGVF----EGELRQHRWMTDRHKTDEYGMVLKQLLEHGLSFED 56
D 1 MLLAIDVRNTHTVVGLSGSKEHAKVQWRIQTESEVTADELALIIDGLI---GDSER 57
QY 57 VKGIIVSSVPPIMFALERCKEYFKIKP-LVVGPGIKTGLNKNYENPREVGADRIYNAV 115
D 58 LAGAAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDPNPKVGVADRIYNAV 117
QY 116 AGHLXGSLPIIIVDFGTATTCTYVINEEKHYMGVITPGIMISAEALYSRAAKLPRIETK 175
D 118 AAYDFRKAIVVDVFGSSICVDVVSAGFELGAIAPGVQVSDAAARSAAARVELAR 177
QY 176 PSSVVGKNTVSAMQSGILYGVQVGEIVKRMKEAKQ-----EPKVIATGGLAKLIS 230
D 178 PRSVGKNTVECMQAGVVFAGLVGRLVGRIRVDVSGFSDHVDVAIVATGHTAPLLPE 237
QY 229 EESNVIDVDPFLTLKGLYMLYERNANLQ 257
D 238 PELHTVDHDLRLTLHGLRLVFERNRAQ 266

RESULT 9
A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70955
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: A70955
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-272 <COL>
A: Cross-references: GB:295557; GB:AL123456; NID: g3242276; PIDN: CAB08944.1; PID: g2113976
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv3600c
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 35.2%; Score 471.5; DB 2; Length 272;
Best Local Similarity 35.2%; Pred. No. 2.9e-30;
Matches 94; Conservative 67; Mismatches 93; Indels 13; Gaps 4;

QY 1 MIFVLDVGNNTNAVIGVF----EGELRQHRWMTDRHKTDEYGMVLKQLLEHGLSFED 56
D 1 MLLAIDVRNTHTVVGLSGSKEHAKVQWRIQTESEVTADELALIIDGLI---GDSER 57
QY 57 VKGIIVSSVPPIMFALERCKEYFKIKP-LVVGPGIKTGLNKNYENPREVGADRIYNAV 115
D 58 LAGAAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDPNPKVGVADRIYNAV 117
QY 116 AGHLXGSLPIIIVDFGTATTCTYVINEEKHYMGVITPGIMISAEALYSRAAKLPRIETK 175
D 118 AAYDFRKAIVVDVFGSSICVDVVSAGFELGAIAPGVQVSDAAARSAAARVELAR 177
QY 176 PSSVVGKNTVSAMQSGILYGVQVGEIVKRMKEAKQ-----EPKVIATGGLAKLIS 230
D 178 PRSVGKNTVECMQAGVVFAGLVGRLVGRIRVDVSGFSDHVDVAIVATGHTAPLLPE 237
QY 231 SNVIDVDPFLTLKGLYMLYERNANLQ 257
D 238 LHTVDHDLRLTLHGLRLVFERNLEVO 264

RESULT 10
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain RI)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: E75516
R: White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A: Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI.
A: Reference number: A75250; MUID: 20036896; PMID: 10567266
A: Accession: E75516
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-262 <WHI>
A: Cross-references: GB:AE001905; GB:AE00513; NID: g6458144; PIDN: AAF10040.1; PID: g645
A: Experimental source: strain RI
C: Genetics:
A: Gene: DR0461
A: Map position: 1
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.2%; Score 405.5; DB 2; Length 262;
Best Local Similarity 31.7%; Pred. No. 5.1e-25;
Matches 84; Conservative 60; Mismatches 104; Indels 17; Gaps 4;

QY 2 IFVLDVGNNTNAVIGVF--EGELRQHRWMTDRHKTDEYGMVLKQLLEHGLSFEDVKGI 60
D 6 LLAVDIGNTTVLGLADASGALHTWRIRTNREMLPDDLALQLHGLFTLAGAIP--RAA 63
QY 61 IVSSVPPIMFALERCKEYFKIKP-LVVGPGIKTGLNKNYENPREVGADRIYNAVAGIHL 120
D 64 VISSVAPPVGENVALAKRHFMDAPVSAENLPDVTVELDTPGSGVADRLCN-----L 117
QY 121 YGSP-----LIIVDFGTATTCTYVINEEKHYMGVITPGIMISAEALYSRAAKLPRIE 172
D 118 FGAEXLYGLDYAVVDFCTSNFVVGRRFLGGLATGATQVSDALFAAAKLPRI 177
QY 173 ITKPSVVGKNTVSAMQSGILYGVQVGEIVKRMKEAKQPKVIATGGLAKLISEEN 232
D 178 LQAPETAIGKNTVHALQSLVFGYAEVMDGLLRRAELPGEAVAVATGFSRTVQGIQ 237
QY 233 VIDVDPFLTLKGLYMLYERNANLQ 257
D 238 EIDYDETTLRLGLVELWASRSEVR 262

RESULT 11
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: D72320
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: D72320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <ARN>
A:Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498141
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0883
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 29.8%; Score 399.5; DB 2; Length 246;
Best Local Similarity 39.2%; Pred. No. 1.4e-24;
Matches 98; Conservative 45; Mismatches 94; Indels 13; Gaps 7;
QY 1 MIFVLDVGNNAVLGVFEGERLQHRMETDRHKTEDEYGMVLVKQLLEHGLSFEDVGI 60
Db 1 MYLLVDVGNTHSVFSITDGTFRWRRLSTGVFTQDELFLSHLPL--GDAMREIRGI 57
QY 61 IVSSVVPPIFMFAERMCKEYFKIKPLVVGPGIKTGL--NIKYPNREVGGADRIVNAVAG 117
Db 58 GVASVVPQNTVIERFSQYFHSIPIW--KAKNCVKNVK--NPSEVGADRVANVAF 113
QY 118 IHLYGSLPIVDFGATTTCYINEEKHYMGVITPGIMISALYRAAKLPRIETKPS 177
Db 114 VREYKNGIITDMGTATVDLV-VNGSYEGGAILPFFGMVHSLFRGTAKLPVLEVKPAD 172
QY 178 SVVGKNTVSAMSGILYGVGQVEGIVKRMKEAKQEPKVIATGGLAKLISESNVIDV 237
Db 173 FVVGKDTENIRLGVVNGSVYALEGIIIGRIKEVYDLP-VLVTGGQSKIVKDMIR-HEIF 230
QY 238 DPFLTLKGLY 247
Db 231 DEDLTIKGY 240
RESULT 12
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C:Accession: D71326
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDowell, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete:
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: D71326
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <COL>
A:Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0431
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 25.5%; Score 342.5; DB 2; Length 273;
Best Local Similarity 30.4%; Pred. No. 5.8e-20;
Matches 77; Conservative 61; Mismatches 104; Indels 11; Gaps 4;
QY 2 IFVLDVGNNAVLGVFEGERLQHRMETDRHKTEDEYGMVLVKQLLEHGLSFE 55
Db 1 MLLIDVGNSHVVGII--QGENGRVCVRELFRLAPDARKTQDEYSLIHALCERAGVGRA 58

QY 56 DVKGIIVSVVPPIMFAERMCKEYFKIKPLVVGPGIKTGLNLIKYNP--REVGDRIYV 113
Db 59 SLRDAFISSVPLVTKTIADAVAQISGVQVVGFWAYEHLVPRIPEVRAEIGTDLVAN 118
QY 114 AVAGIHLVGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISALYRAAKLPRIE 173
Db 119 AVAAVYVHFSAVWVDCGTFALTAVDGTGLIGQVAIAFGLRATVOSLHTGTGAQLPLVPL 178
QY 174 TKPSSVVGKNTVSAMSGILYGVGQVEGIVKRMKEAKQEPKVIATGGLAKLISESNV 233
Db 179 ALPDSVLGKDTTHAVQGVVVRGTLFVIRAMIAQCQKELCGRCAAVITGSLRFLSSEVD- 237
QY 234 IDVDPFLTKGL 246
Db 238 FPPIDAQLTSLGL 250
RESULT 13
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: F70165
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whiston, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Venter, J.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70165
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <KLE>
A:Cross-references: GB:AE001154; GB:AE000783; NID:g2588431; PIDN:AAC66882.1; PID:g268
A:Experimental source: strain B31
Query Match 20.6%; Score 276.5; DB 2; Length 262;
Best Local Similarity 27.8%; Pred. No. 1e-14;
Matches 71; Conservative 63; Mismatches 100; Indels 21; Gaps 7;
QY 4 VLDVGNNAVLGVFEGERLQHRMETDRHKTEDEYGMVLVKQLLEHGLSFEVDKGIIVS 63
Db 10 IIDIGNTSTAFALQDNQVNLFIKMTNLMRYDE---VYSFFE-ENEDF-NVNVKVIS 63
QY 64 SVVPPIMFAERMCKEYFKIKPLVVGPGIKTGLNIKYE---NPRE-----VGADRIYNAV 115
Db 64 SVVPIINETFKNVIFSFYFKIKPLFI-----GFDLNYDLTFNPKYKSKFLGSDVFNALV 117
QY 116 AGIHLVGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISALYRAAKLPRIET 174
Db 118 AAIENYSPFNVLVDLGTACTIFAVSRQDGIIGIINSPLINFNSLLDNAYLIKFPIS 177
QY 175 KPSSVVGKNTVSAMSGILYGVGQVEGIVKRMKEAKQEPKVIATGGLAKLISESNV 234
Db 178 TPNLLERTTSVNSGLFYQYKLYEGVYRDIKQYKKKFNLIITGGNADLILSLIEE 237
QY 235 DVDPFLTKGLYML 249
Db 238 FIFNIHLTVEGVRI 252
RESULT 14
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: B81009
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigani, V.; Pizsa, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.27037 Seconds
(without alignments)
2061.866 Million cell updates/sec

Title: US-09-813-453A-45
Perfect score: 1341
Sequence: 1 MIFVLDVGNVTNAVGLVFEEG.....LKGLYMLYERNANLQHEKGE 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	901	67.2	233	1	YACB_BACSU	P37564 bacillus su
2	138	10.3	267	1	BAF_BORPE	Q45338 bordetella
3	106	7.9	295	1	Y188_CLOPE	P26832 clostridium
4	103.5	7.7	736	1	EF2_AERPE	Q9yc19 aeropyrum p
5	96	7.2	933	1	CS54_YEAST	P30665 saccharomyc
6	96	7.2	1067	1	CABP_CLOPE	Q8xbh3 clostridium
7	93	6.9	326	1	LPXD_AQUAE	O66817 aquifex aeo
8	92.5	6.9	361	1	MPGL_YEAST	P41940 saccharomyc
9	91.5	6.8	291	1	PYRD_SULAC	O08358 sulfolobus
10	90	6.7	428	1	PURA_CLOAB	Q97d87 clostridium
11	89.5	6.7	256	1	H161_PSEAE	Q9hu44 pseudomonas
12	89.5	6.7	571	1	PT1_BUCAI	O9exi6 buchnera ap
13	89	6.6	304	1	HEN3_AQUAE	O66621 aquifex aeo
14	88.5	6.6	355	1	BUK_CLOBE	Q03619 clostridium
15	88.5	6.6	793	1	DCMA_METSO	P26692 methanothi
16	88	6.6	597	1	IF2P_METAC	Q8tql5 methanosarc
17	88	6.6	1432	1	SK13_YEAST	P17883 saccharomyc
18	87.5	6.5	331	1	FABH_HELPY	O24994 helicobacte
19	87	6.5	415	1	SAHH_SULTO	Q975t0 sulfolobus
20	86	6.4	560	1	ARAB_BACSU	P45424 bacillus su
21	86	6.4	1325	1	MRP4_HUMAN	O15439 homo sapien
22	85.5	6.4	331	1	FABH_HELPJ	Q92mn0 helicobacte
23	85.5	6.4	1065	1	CARB_BACCL	P46537 bacillus ca
24	84.5	6.3	501	1	DLDH_PEA	P31023 pisum sativ
25	84.5	6.3	733	1	FLHA_HELPY	Q9zm40 helicobacte
26	84.5	6.3	733	1	FLHA_HELPY	O06758 helicobacte
27	83	6.2	679	1	GR78_KLULA	P22010 kluyveromyc
28	83	6.2	788	1	PUR2_YARLI	Q99148 y bifunctio
29	83	6.2	848	1	CLPC_MYCTU	O06286 mycobacteri
30	82.5	6.2	613	1	YGAR_YEAST	P50080 saccharomyc
31	82.5	6.2	1057	1	CARB_STAAN	P58940 staphylococ
32	82.5	6.2	1064	1	CARB_BACST	O50302 bacillus st
33	82.5	6.2	1088	1	DP0D_SOYBN	O48901 glycine max

34	82.5	6.2	1197	1	DPOM_PODAN	Q01529 podospora a
35	82.5	6.2	2201	1	TENA_HUMAN	P24821 homo sapien
36	81.5	6.1	541	1	YQEB_ECOLI	Q46808 escherichia
37	81.5	6.1	606	1	EDD_RHIME	Q92350 rhizobium m
38	81.5	6.1	1146	1	MMIC_MYCTU	Q50385 mycobacteri
39	81	6.0	176	1	IPYC_GLUOX	O05545 gluconobact
40	81	6.0	473	1	AK_METUA	Q57991 methanococc
41	81	6.0	499	1	DLDH_YEAST	P09624 saccharomyc
42	81	6.0	675	1	CMC2_HUMAN	Q9ui80 homo sapien
43	80.5	6.0	328	1	GLK_STAXY	Q56198 staphylococ
44	80.5	6.0	401	1	CHRA_ALCEU	P17551 alcaligenes
45	80.5	6.0	424	1	SAHH_METKA	P58855 methanopyru

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564:
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chol S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Fabret C., Ferrari E., Foulger D.,
RA Entian K.D., Errington J., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medique C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosati V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D26185; BAA05305.1; -
CC EMBL: 299104; CAB11846.1; -
CC Subtilisin; BG10133; yacB.
CC InterPro: IPR004619; Baf.
CC Pfam: PF03309; Bvg_acc_factor; 1.
CC TIGRFAMs: TIGR00671; baf; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 233 AA; 26217 MW; AA956E732C15DF44 CRC64;
Query Match 67.2%; Score 901; DB 1; Length 233;
Best Local Similarity 75.5%; Pred. No. 2.2e-65;
Matches 163; Conservative 29; Mismatches 24; Indels 0; Gaps 0;
1 MIFVLVGVNTNVLGVFESEGLRQHWRTDHRKTEDEYMLVKLLEHGLSFEDYVKGI 60
1 MLIVDVGNTNVLGVTHDGLKLEHYHRIETSRHKTDEFGMLIRSLFDHSGLMFEQIDGI 60
61 IVSSVVPPIFALERMCEYFKIKPLVVGPGIKTGLNIRYENPREVGADRIYNAVAGIHL 120
61 IISVVVPIFALERMCTYFHEIPOIVGPGMKTGLNIRYDNEKVEGADRIYNAVAGIHL 120
121 YGSLPIVDFGATTTCYINEEKHYMGVITPGIMTSAEALYSRAAKLPRIETKPSVV 180
121 YGNPLVDFGATTTCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIETRPDNI 180
181 GKNTVSAMSGILVGV 216
181 GKNTVSAMSGILVGV 216
RESULT 2
BAF_BORPE STANDARD; PRT; 267 AA.
ID Q45373;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Bordetella.
NCBI_TaxID=520;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=BP504;
RC Wood G.E., Friedman R.L.;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshazer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli".
RL J. Bacteriol. 177:3801-3807(1995).
RN [2]
RN SEQUENCE OF 1-38 FROM N.A.
RP STRAIN=BP504;
RC Wood G.E., Friedman R.L.;
RT "Identification of a bira homolog in Bordetella pertussis".
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 239-267 FROM N.A.
RP STRAIN=BP536;
RC Allen A.G., Maskell D.J.;
RX MEDLINE=96419162; PubMed=8821935;
RA "The identification, cloning and mutagenesis of a genetic locus
RT

RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis".
RL Mol. Microbiol. 19:37-52(1996).
CC 1- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BVCAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U12020; AAA75361.1; -
CC EMBL: AF016461; AAC68834.1; -
CC EMBL: X90711; CAA62242.1; -
CC InterPro: IPR004619; Baf.
CC Pfam: PF03309; Bvg_acc_factor; 1.
CC Transcription regulation; Activator.
CC SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;
Query Match 10.3%; Score 138; DB 1; Length 267;
Best Local Similarity 23.8%; Pred. No. 0.00045;
Matches 63; Conservative 46; Mismatches 96; Indels 60; Gaps 12;
1 MIFVLVGVNTNVLGVFE-----EGELRQHWRTDHRKTEDEYVGL 42
1 MLIVDVGNTNVLGVTHDGLKLEHYHRIETSRHKTDEFGMLIRSLFDHSGLMFEQIDGI 59
43 VKQLLEHGLSFEDYVKGIIVSSVVPPIFALERMCE-KYFKIKPLVVGPGIKTGLNIRY 101
60 VAGLRGEAIAATLRAG-----GCDIRWLRAQPLAM--GLRNG----YR 97
102 NPREVGADR---IVNAVAGIHLVGSPLIIVDFGATTTCYINEEKHYMGVITPGIMISA 158
98 NPDQLGADRWACMVGVLARQPSVHPPLVASFATLTDTIGPDNVPFGGLILPGPAMMR 157
159 EALYSRAAKLPRIETKPSVVGVK---NTVSAMSGILVGVGVGVGVGVGVGVGVGVGV 212
158 GALAYGTAHLPLAD-----GLVADYPIDTHQAIASGIA---AAQAGAIYVQWLAGRQY 209
213 QEPKV-IATGGLAKLISEESNIDV 236
210 QAPEIVVAGGWNPEVQEAERLLAV 234
RESULT 3
Y188_CLOPE STANDARD; PRT; 295 AA.
ID Y188_CLOPE
AC P26832;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein CPE0188.
GN CPE0188.
OS Clostridium perfringens.
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
NCBI_TaxID=1502;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yanashita A.,
RT Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater".
RL Proc. Natl. Acad. Sci. U.S.A. 99:9996-1001(2002).
RN [2]
RN SEQUENCE OF 114-295 FROM N.A.
RP STRAIN=CPN50;
RC

FT DOMAIN 550 932 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 933 1067 ALLOSTERIC DOMAIN.
FT REPEAT 1 549
FT REPEAT 550 1067
FT NP_BIND 133 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 298 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 303 MANGANESE 2 (BY SIMILARITY).
FT METAL 823 823 MANGANESE 3 (BY SIMILARITY).
FT METAL 835 835 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1067 AA; 118897 MW; A4C2E518224611A6 CRC64;
Query Match 7.2%; Score 96; DB 1; Length 1067;
Best Local Similarity 24.4%; Pred. No. 5.3;
Matches 66; Conservative 37; Mismatches 82; Indels 86; Gaps 14;
QY 37 DEYGMVLVQLLEHGLS-FEDVKGIIVSVVPPIMPALER-----MCEK 79
b 748 DRY--LMGREIEVDASDCEDY-----LVPGIMEHLERAGVHSGDSITMYPQAQNSDK 798
QY 80 YFK-----IKPLVVGPGIKTGNIKY-----ENPREVGADRIVNAVAGIHLVGS 123
Db 799 IKEDVLDYTKKALSIGIKGMINIOFIEPEGKLYIEVNPRASTRVPIYISKVSGVP---- 854
QY 124 PLIIVDFGPTATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETIKPSSV---- 179
Db 855 ---IVDIATRIMLGEKLDLGTGTGYKPELVSKVPVFSQKLPNVEVSLGPENRSTG 911
QY 180 ----VGKNTVSAQSGILYV--YVQG-----VEGIVKRMKE-----AKQEPKV----IAT 220
Db 912 EYLVGVRNVFEALYKGFVGSMTYTDGKTKILATIKKHDKKFMELAKDLKLYNFYIAT 971
QY 221 GGLAK-----LISEES-NVIDV 237
Db 972 TGTAKELREAGIDAKEVRRIGBESPNIMDLI 1002
RESULT 7
LPXD_AQUAE
ID LPXD_AQUAE STANDARD; PRT; 326 AA.
AC O66817;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
(EC 2.3.1.1).
LN LPXD OR AQ_538.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: UDP-3-O-(3-hydroxytetradecanoyl)glucosamine +
(R)-3-hydroxytetradecanoyl [acyl-carrier protein] -> UDP-2,3-bis(3-
hydroxytetradecanoyl)glucosamine + [acyl-carrier protein].
CC -1- PATHWAY: Lipid A biosynthesis; third step.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERASE HEXAPEPTIDE REPEAT FAMILY.
LPXD SUBFAMILY.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
frameshift in position 105.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.
CC
DR EMBL; AE000694; AAC06767.1; ALT_FRAME.
DR HSP; P10440; 1LXA.
DR InterPro: IPR001451; Hexapep_transf.
DR Pfam: PF00132; hexapep; 9.
DR PROSITE: PS00101; HEXAPEP_TRANSFERASES; FALSE_NEG.
KW Transferase; Acyltransferase; Lipid A biosynthesis; Lipid synthesis;
KW Repeat; Complete proteome.
SQ SEQUENCE 326 AA; 35800 MW; BF785A3ADB9398C8 CRC64;
Query Match 6.9%; Score 93; DB 1; Length 326;
Best Local Similarity 26.3%; Pred. No. 2.3;
Matches 47; Conservative 24; Mismatches 62; Indels 46; Gaps 9;
QY 86 LVVVGPIKTLGNIKYENPREVGADRIV--NAV--AGIHLV-----GSPLIIV 128
Db 116 VVIGKNVIGRNKVIPTFTYVGDNTVIGDNTVIFSGVHIYRNTVIGRNVRIHSGAVIGAD 175
QY 129 DFGTATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETIKPSSVGNKT---- 184
Db 176 GFGYHTQBGIKKIPHIGGVIIDNVEIGANTTIDRAL-----IE-----NTLIGKNTKIDN 227
QY 185 -----VSAMSGILYGVYGVQGVIGVIRKMKBEAKQPKVIATG--GLAKLISEESNVI 234
Db 228 LVNVAHNCKVGNENILVSVQV--LSGSVKTK-----NVILAGQGVADHVEIGDNVI 278
RESULT 8
MPGL_YEAST
ID MPGL_YEAST STANDARD; PRT; 361 AA.
AC P41940;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Mannose-1-phosphate guanylyltransferase (EC 2.7.7.13) (ATP-mannose-1-
phosphate guanylyltransferase) (NDP-hexose pyrophosphorylase).
GN MPGL OR PSA1 OR YDL055C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96418871; PubMed=8821656;
RA Benton B.K., Plump S.D., Roos J., Lennarz W.J., Cross F.R.;
RT "Over-expression of S. cerevisiae Gl cyclins restores the viability
of alg1 N-glycosylation mutants.";
RL Curr. Genet. 29:106-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Schultz J., Sprague G.F. Jr.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Bloecker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL PROTEIN WHICH MAY BE INVOLVED IN THE
REGULATION OF CELL CYCLE PROGRESSION.
CC -1- CATALYTIC ACTIVITY: GTP + alpha-D-mannose 1-phosphate ->
diphosphate + GDP-mannose.
CC -1- SIMILARITY: TO S.TYPHIMURIUM CDP-GLUCOSE PYROPHOSPHORYLASE (RFBF).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)


```

CC EMBL: AE007855; AAK81516.1;
DR InterPro: IPR001114; Asucc_synthase.
DR Pfam: PF00709; Adenylsucc_synth: 1.
DR ProDom: PD001188; Asucc_synthase; 1.
DR TIGRFAMs: TIGR00184; purA; 1.
DR PROSITE: PS01266; ADENYLOSUCCLIN_SYN_1; 1.
DR PROSITE: PS00513; ADENYLOSUCCLIN_SYN_2; 1.
KW Purine biosynthesis; Ligase; GTP-binding; Complete proteome.
FT NP_BIND 12 18
FT ACT_SITE 141 141 BY SIMILARITY.
FT ACT_SITE 148 148 BY SIMILARITY.
SQ SEQUENCE 428 AA; 46968 MW; 4248E7A476C5361B CRC64;

Query Match
Best Local Similarity 22.8%; Score 90; DB 1; Length 428;
Matches 49; Conservative 31; Mismatches 69; Indels 66; Gaps 11;

6 DVGNTNAVGVFEGELRQHWMTDRHKTDEYGMVLKQLLEHGL-----SPFDVKGI 60
126 DIGTGGKIG-----PCYTDK---AERSGIRVCDLLHPEVFEKLNLEIKNA 171

QY 61 IVSSVVPIMPALERMCEKYP---KIP-----LVVPGIKTGLNLYENPREVGADR 110
172 IITKYGMDAFDYNEICEYLAFGKELKPFVKDTSVIYVNEIKNGKKVLF-----GAQ- 225

QY 111 IVNAVGHLYGSPLIIVDFGTATTCYVINEEKHYMGVITPGMIS-----AEA 160
226 -----GNLLDIDYG---TYPYVT-SSNTIGGVCVPGAIGTMTITSVAGIAGA 269

QY 161 LYSRAAKLPRIETKPPSSVVGKNTVSAMQSGILYG 195
270 YTTRVGKGP-----FPTELLEDMDGRIREAGFEYG 299

Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
NCBI_TaxID=287;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 15692 / PAOI;
MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Stover C.K., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RA opportunistic pathogen."
RA Nature 406:959-964(2000).
CC -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-
CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -!- PATHWAY: Histidine biosynthesis; fifth step.
CC -!- SUBUNIT: Heterodimer of hisH and hisF (By similarity).

```

```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE004927; AAG08525.1; -.
CC InterPro: IPR003009; FMN_enzyme.
CC InterPro: IPR000570; His_biosynth.
CC Pfam: PF00977; His_biosynth: 1.
CC TIGRFAMs: TIGR00735; hisF; 1.
KW Histidine biosynthesis; Lyase; Complete proteome.
FT ACT_SITE 12 12 POTENTIAL.
FT ACT_SITE 131 131 POTENTIAL.
SQ SEQUENCE 256 AA; 27131 MW; 6486A76CD308FAA7 CRC64;

Query Match
Best Local Similarity 6.7%; Score 89.5; DB 1; Length 256;
Matches 63; Conservative 34; Mismatches 68; Indels 95; Gaps 17;

5 LDVGNNAVGVFEG-----ELRQHWMTDRHKTDEYGMVLKQLLEHGLSPED 56
11 LDVGNRVGVKVFENIRDAGDPVEIARY-----DEQGA-----DRTFLD 52

QY 57 VKGIIVSSVVPIMPALERMCEKYPKIRPLVVGIGIKTGLNLYENPREVGADRI-VNAV 115
53 ITASVDGR--DITLHTVERMASQVF--IPLTVGGGVRSVQDIR--NLLNAGADKVSINTA 106

QY 116 A-----GIHLYGSPLIIVDFGTATTCYVINEEKHYMGVITPGMISAEALYSRAA 166
107 AVFNPEFYGEAADRFSGQCIW-----AIDAKK-----VSAPG----- 139

QY 167 KLPRIE-----TKPSSV-----VGKNTVSAM-QSGILYGY-VGQVEIGVKR 206
140 EAPWEIETHGGRKPTGLDAVLWAKKMEDLGNAGEILLISMDQDGKVGSGYDLG-----VTR 194

QY 207 MKEEAKQEPKVIATGGLAKL 226
195 AISEAVNVP-VIASGGVGNL 213

RESULT 12
PFI_BUCAI
ID PFI_BUCAI STANDARD; PRT; 571 AA.
AC Q9WXI6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9)
DE PTSI OR BU064.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RA Matsumoto K., Morioka M., Ishikawa H.;
RA "Buchnera sp. DNA for ptsH-ptsI-crr operon.";
RA Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RL [2]
RC STRAIN=Tokyo 1998;
RC SEQUENCE FROM N.A.
RX MEDLINE=20445173; PubMed=10993077;
RX Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RT Nature 407:81-86(2000).

```

```

146 ---GNVDTRMRRLUK-BGLYDAVILAYAGV-KRMGYESEITEVLEDFIPAVG 191

RESULT 14
BUK_CLOBE
ID      BUK_CLOBE      STANDARD;      PRT;      355 AA.
DT      AC
005619;
01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Butyrate kinase (EC 2.7.2.7) (BK).
DE      GUK.
OS      Clostridium beijerinckii (Clostridium MP).

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 23.0467 seconds
(without alignments)
2342.388 Million cell updates/sec

Title: US-09-813-453a-45
Perfect score: 1341
Sequence: 1 MIFVLDVGNNTNAVGLVFEEG.....LKGLMYLERNALQHEKGE 262

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 205047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1054	78.6	258	Q9F985	Q9F985 bacillus st
2	960	71.6	254	Q9KGH5	Q9KGH5 bacillus ha
3	872	65.0	259	Q9YAC5	Q9YAC5 listeria mo
4	867	64.7	259	Q92F54	Q92F54 listeria in
5	804	60.0	255	Q9R7M2	Q9R7M2 thermoanaer
6	804	60.0	273	Q97EB4	Q97EB4 clostridium
7	780	58.2	259	Q9XHL5	Q9XHL5 clostridium
8	638	47.6	265	Q9X8N6	Q9X8N6 streptomyce
9	566	42.2	261	Q9A621	Q9A621 caulobacter
10	492.5	36.7	274	Q9CD56	Q9CD56 mycobacteri
11	478	35.6	256	Q9RFE4	Q9RFE4 fusobacteri
12	471.5	35.2	272	Q06282	Q06282 mycobacteri
13	405.5	30.2	262	Q9RX54	Q9RX54 delinococcus
14	399.5	29.8	246	Q9WZY5	Q9WZY5 thermotoga
15	379.5	28.3	212	Q32514	Q32514 desulfovibr
16	342.5	25.5	273	Q83446	Q83446 treponema p

17	276.5	20.6	262	16	051477	O51477 borrelia bu
18	169.5	12.6	592	16	Q9JXF1	Q9JXF1 neisseria m
19	168	12.5	592	16	Q9JW17	Q9JW17 neisseria m
20	167.5	12.5	229	16	O67753	O67753 aquifex aeo
21	159	11.9	56	2	P94305	P94305 bacillus ps
22	158	11.8	295	16	Q8Y2M4	Q8Y2M4 raietonia s
23	156.5	11.7	257	16	P74045	P74045 synechocyst
24	147	11.0	276	16	Q8YQD7	Q8YQD7 anabaena sp
25	136	10.1	248	16	Q9HWC1	Q9HWC1 pseudomonas
26	121	9.0	224	16	Q98O93	Q98O93 mycoplasma
27	117	8.7	242	16	Q9PC14	Q9PC14 xylella fas
28	113	8.4	209	16	Q9PIA9	Q9PIA9 campylobact
29	112.5	8.4	223	16	Q9ZKY6	Q9ZKY6 helicobacte
30	112.5	8.4	324	17	Q97Y22	Q97Y22 sulfolobus
31	105.5	7.9	312	16	Q8R8N4	Q8R8N4 thermoanaer
32	103.5	7.7	315	16	Q8RDE9	Q8RDE9 thermoanaer
33	103.5	7.7	351	10	Q9M0A3	Q9M0A3 arabidopsis
34	103.5	7.7	597	17	Q30225	Q30225 archaeoglob
35	100.5	7.5	480	17	Q97CH2	Q97CH2 thermoplasm
36	98.5	7.3	223	16	Q25533	Q25533 helicobacte
37	97.5	7.3	400	16	Q9PGS8	Q9PGS8 xylella fas
38	97.5	7.3	816	16	Q8RB00	Q8RB00 thermoanaer
39	96.5	7.2	828	16	Q9PBB9	Q9PBB9 xylella fas
40	96	7.2	1067	16	Q8XHB3	Q8XHB3 clostridium
41	95	7.1	1026	16	Q8X299	Q8X299 escherichia
42	94.5	7.0	400	17	Q96Y60	Q96Y60 sulfolobus
43	93	6.9	607	17	Q976F4	Q976F4 sulfolobus
44	92	6.9	317	17	P95892	P95892 sulfolobus
45	92	6.9	592	17	O30212	O30212 archaeoglob

ALIGNMENTS

RESULT 1

Q9F985 PRELIMINARY; PRT; 258 AA.
ID Q9F985
AC Q9F985;
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Putative 32 kDa replication protein.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V;
RA Vasquez C., Pichuanes S., Saavedra C.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF198621; AG28531.1;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match	78.6%	Score 1054;	DB 2;	Length 258;
Best Local Similarity	78.0%;	Pred. No. 7.1e-83;		
Matches 199;	Conservative 23;	Mismatches 33;	Indels 0;	Gaps 0;
QY	1	MIFVLDVGNNTNAVGLVFEEGELRQHRMETDRHKTEDEYGLVYKQLEHEGLSFEVDKGI	60	
Db	1	MIFVLDVGNNTNAVGLVYDGDGLKHHWRIETSRGKTEDEYGMTIKALLNHVGLQFSDIDGI	60	
QY	61	IVSSVVPPIFMFALERCERYFKIKPLVVGPGIKTGLNIKENPREYGADRIYNAVAGIHL	120	
Db	61	IISSVVPPIFMFALERCERYFKIKPLVVGPGIKTGLNIKENPREYGADRIYNAVAGIHL	120	
QY	121	YGSPLIIVDFGATTTCYCYNEEKHYNGVITPGIMISAEALYSRAAKLPRIETKPSVV	180	
Db	121	YGSPLIIVDFGATTTCYCYNEEKHYNGVITPGIMISAEALYSRAAKLPRIETKPSVV	180	

```
QY 181 GKNVTSAMQSGILGYVGQVEGIVKRMKEAKQEPKVIATGGTGLAKLISESNVIDVDPF 240
DB 181 GKNVTSAMQSGILGYVGQVEGIVKRMKEAKSPVPKVIATGGTGLAKLISESNVIDVDPF 240
QY 241 LTLKGLYMLYERNAN 255
DB 241 LTLTGLKILYKKNVD 255

RESULT 2
Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
NCBI_TaxID=86665;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001507; BAB03805.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRfams; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 71.6%; Score 960; DB 16; Length 254;
Best Local Similarity 69.2%; Pred. No. 8.9e-75;
Matches 175; Conservative 35; Mismatches 43; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTNAVGLVGFEEGELRQHRMETDRHKTDEYGMVQLLEHGLSFEDVKGI 60
DB 1 MILVIDVGNNTVGVYQDETLVHWRLATSRQKTEDEYAMTVRSFLDHAGLQFDIDGI 60
QY 61 IVSSVVPPIAFALERMCEKFKIKPLVGVGPKITGLNIKYENPREVGADRVNAVAGIHL 120
DB 61 VISSVVPPIAFALERMCEKFKIKPLVGVGPKITGLNIKYENPREVGADRVNAVAGIHL 120
QY 121 YGSLPIVDGTTATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETTKPSSV 180
DB 121 YGPAIVVDGTTATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETAKPQVV 180
QY 181 GKNVTSAMQSGILGYVGQVEGIVKRMKEAKQEPKVIATGGTGLAKLISESNVIDVDPF 240
DB 181 GTNTIDSMQSGIFGYVQVGVKRMKAQAEKPKVIATGGTGLAKLIGTSETIDVDSF 240
QY 241 LTLKGLYMLYERN 253
DB 241 LTLKGLQIYKKN 253

- RESULT 3
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.
```

```
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
NCBI_TaxID=1639;
[1]
SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-Del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novellia S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1; -
DR MEROPS; M41.009; -
DR Listeriust; LMO00221; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRfams; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F108E049D80025B7 CRC64;

Query Match 65.0%; Score 872; DB 16; Length 259;
Best Local Similarity 61.7%; Pred. No. 3.5e-67;
Matches 156; Conservative 49; Mismatches 48; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTNAVGLVGFEEGELRQHRMETDRHKTDEYGMVQLLEHGLSFEDVKGI 60
DB 1 MILVIDVGNNTVGVYQDETLVHWRLATSRQKTEDEYAMTVRSFLDHAGLQFDIDGI 60
QY 61 IVSSVVPPIAFALERMCEKFKIKPLVGVGPKITGLNIKYENPREVGADRVNAVAGIHL 120
DB 61 VISSVVPPIAFALERMCEKFKIKPLVGVGPKITGLNIKYENPREVGADRVNAVAGIHL 120
QY 121 YGSLPIVDGTTATTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETTKPSSV 180
DB 121 YGTPVIVDGTATTCYIDESGVYQGAIPGIMISTEALYNRAAKLPVDAESSQII 180
QY 181 GKNVTSAMQSGILGYVGQVEGIVKRMKEAKQEPKVIATGGTGLAKLISESNVIDVDPF 240
DB 181 GKSTVSSMQAGIFGVGQCEGIIAEMKKQSNASPVVATGGTGLARMITKSSAVDIDPF 240
QY 241 LTLKGLYMLYERN 253
DB 241 LTLKGLLELYRN 253

RESULT 4
Q92F54 PRELIMINARY; PRT; 259 AA.
AC Q92F54;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lin0253.
GN LIN0253.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
NCBI_TaxID=1642;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
```

```

RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domani E., Dominguez-Bernal G., Duchaud E., Duran L., Dussurget O.,
RA Entian K.-D., Fishi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkak G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wenland J., Cossart P.,
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95486.1; -.
DR ListiList; LIN00253; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 64.7%; Score 867; DB 16; Length 259;
Best Local Similarity 60.9%; Pred. No. 9.5e-67;
Matches 154; Conservative 51; Mismatches 48; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTAVLGVFEGLRQHWMETDRHKTDEYGLVKQLLEHGLSFEDVKGI 60
Db 1 MIFVLDVGNNTAVLGVFEGLRQHWMTDRHKTDEYGLVKQLLEHGLSFEDVKGI 60
QY 61 IVSSVPPIMFALERCKEYFKIKPLVGVPGIKTGLNKKYENPREVGADRIYNAVAGIHL 120
Db 61 IVSSVPPIMFALERCKEYFKIKPLVGVPGIKTGLNKKYENPREVGADRIYNAVAGIHL 120
QY 121 YGSLPIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 121 YGSLPIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
QY 181 KNTVSMQSGILYGVQGVGIVKRMKEEAQKPKVIATGGLAKLISESNVIDVDPF 240
Db 181 KNTVSMQSGILYGVQGVGIVKRMKEEAQKPKVIATGGLAKLISESNVIDVDPF 240
QY 241 LTKGLYMLYERN 253
Db 241 LTKGLYMLYERN 253

RESULT 5
QY Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory factor.
DE TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013180; BAM25520.1; -.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;

Query Match 60.0%; Score 804; DB 16; Length 255;

```

```

Best Local Similarity 58.5%; Pred. No. 2.5e-61;
Matches 148; Conservative 51; Mismatches 54; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTAVLGVFEGLRQHWMETDRHKTDEYGLVKQLLEHGLSFEDVKGI 60
Db 1 MIFVLDVGNNTAVLGVFEGLRQHWMTDRHKTDEYGLVKQLLEHGLSFEDVKGI 60
QY 61 IVSSVPPIMFALERCKEYFKIKPLVGVPGIKTGLNKKYENPREVGADRIYNAVAGIHL 120
Db 61 IVSSVPPIMFALERCKEYFKIKPLVGVPGIKTGLNKKYENPREVGADRIYNAVAGIHL 120
QY 121 YGSLPIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 121 YGSLPIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
QY 181 KNTVSMQSGILYGVQGVGIVKRMKEEAQKPKVIATGGLAKLISESNVIDVDPF 240
Db 181 KNTVSMQSGILYGVQGVGIVKRMKEEAQKPKVIATGGLAKLISESNVIDVDPF 240
QY 241 LTKGLYMLYERN 253
Db 241 LTKGLYMLYERN 253

RESULT 6
QY Q8R7E4 PRELIMINARY; PRT; 273 AA.
AC Q8R7E4;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory factor.
DE CAC3200.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibscn R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007815; RAK81136.1; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;

Query Match 60.0%; Score 804; DB 16; Length 273;
Best Local Similarity 59.1%; Pred. No. 2.8e-61;
Matches 152; Conservative 49; Mismatches 52; Indels 4; Gaps 1;

QY 1 MIFVLDVGNNTAVLGVFEGLRQHWMETDRHKTDEYGLVKQLLEHGLSFEDVKGI 60
Db 12 VILVLDVGNNTAVLGIYNDTKLTAEMRLSTDLRSADYGIQVMNLFQDKLPTLVEGV 71
QY 61 IVSSVPPIMFALERCKEYFKIKPLVGVPGIKTGLNKKYENPREVGADRIYNAVAGIHL 120
Db 72 IISVVPNIMYSLHEMIRKIFKINPLVGVPGIKTGLNKKYENPREVGADRIYNAVAGIHL 131
QY 121 YGSLPIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
Db 132 YKRSLLIIDFGTATTFCVAVRENGDYLGAICPGKVSSEALFEKAALKPRVELIKPAYAI 191

```

QY 181 GNTVSMQSGILYGVQGVIGVIRKMK-----BEAKQPKVIATGGLAKLISESNVIDV 236
 DB 192 CKNTISSIQSGIVYIGVQVYIVIRKMKELQEGEKEPLVATGGLAKLISEAKNVDV 251
 QY 237 VDPFLTKGLYMLYERN 253
 DB 252 INPFLTLEGLRIIYERN 268

RESULT 7

Q8XHL5 ID Q8XHL5 PRELIMINARY; PRT; 259 AA.
 AC Q8XHL5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein CPE2468.
 GN CPE2468.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-13 / TYPE A;
 RX PubMed-11792842;
 RA Shiba T., Ogasawara N., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL: AP003194; BAB82174.1; -;
 DR InterPro: IPR004619; Baf.
 DR InterPro: IPR000515; BPD.transp.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMS: TIGR00671; Baf; 1.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

Query Match 58.2%; Score 780; DB 16; Length 259;
 Best Local Similarity 60.3%; Pred. No. 3e-59;
 Matches 155; Conservative 42; Mismatches 56; Indels 4; Gaps 1;

QY 1 MIFVLDVGNNTNAVIGVFEEGELRQHWRMETDRHKTDEYGMVYKQLEHGLSFEDVKGI 60
 DB 1 MILLIDVGNNTNVLGLTHDNEKYATSWRISTDSKTSDEYSIQVQWLFNQAALNPEDVEGI 60
 61 IVSSVVPPIFMALERMCEKFKIKPLVVGPGIKTGLNIKYENPREVGADRIINAVAGIHL 120
 61 IISVVVPMHSLNNVVRKCFCKEPIVVGPGIKTGINIKYDNKPNKVGADRIINAVAAFEK 120
 QY 121 YGSPLIIVDGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPSVV 180
 DB 121 HKPMIILIDGTATTCATKEDGYLGNPCGQIQISADALFRAAKLPRIELKPSVI 180
 QY 181 GNTVSMQSGILYGVQGVIGVIRKMKKE- ---AKQPKVIATGGLAKLISESNVIDV 236
 DB 181 CKNTVSMQAGIIVYIGVYIVIRKMKEMMDLGEKEPFLVATGGLAKLIVSETVIDE 240
 QY 237 VDPFLTKGLYMLYERN 253
 DB 241 VDRKLTLEGLKILYERN 257

RESULT 8

Q9X8N6 ID Q9X8N6 PRELIMINARY; PRT; 265 AA.
 AC Q9X8N6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical protein SCO3380.
 GN SCO3380 OR SCE94.31C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL: AL049628; CAB40880.1; -;
 DR InterPro: IPR004619; Baf.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMS: TIGR00671; Baf; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 47.6%; Score 638; DB 16; Length 265;
 Best Local Similarity 47.9%; Pred. No. 5.5e-47;
 Matches 126; Conservative 55; Mismatches 74; Indels 8; Gaps 3;

QY 1 MIFVLDVGNNTNAVIGVFEEGELRQHWRMETDRHKTDEYGMVYKQLEHGLSFED--- 56
 DB 1 MLTIDVGNTHVTLVGLFDGEDIVHWRISTDSRTADELAVLQGLMGHPLGLGDLGDC 60
 QY 57 VKGIIVSSVVPPIFMALERMCEKFKIKPLV-VGPGIKTGLNIKYENPREVGADRIINAV 115
 DB 61 IDGIAICATNPVSVLHELREVTTRYGVDPVAVLVEPGVKTGVPILTDPKEVGADRIINAV 120
 QY 116 AGIHLGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETK 175
 DB 121 AAVELYGPAIVVDFGTATTTFDAVSARGEYIGGVIAPIEISVEALGVGAQURKIEVAR 180
 QY 176 PSSVVGKNTVSMQSGILYGVQGVIGVIRKMKKEAKQEP- ---KVATGGLAKLISESN 232
 DB 181 PRSVIGKNTVEAMQSGIVYIGVYIVIRKMKEMMDLGEKEPFLVATGGLAPWVLGESS 240
 QY 233 VIDVDPFLTKGLYMLYERNAN 255
 DB 241 VIDEHEPWLTLMLGLRLVYERNVS 263

RESULT 9

Q9A6Z1 PRELIMINARY; PRT; 261 AA.

ID Q9A6Z1
AC Q9A6Z1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.

OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldbyud T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
Fotocika I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Hart D.H.,
Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
Rutterback T., Tran K., Wolf A., Vamathevan J., Emmolaeva M., White O.,
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RL EMBL; AE005867; AKK23910.1; -;
DR TIGR; CC1935; -;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvq_acc.factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19EG0D7B0714EF5 CRC64;

Query Match 42.2%; Score 566; DB 16; Length 261;
Best Local Similarity 42.0%; Pred. No. 8.8e-41;
Matches 107; Conservative 55; Mismatches 91; Indels 2; Gaps 1

QY 1 MIFVLDVGNTNAVGLGVFEELGKLPVVGGIKTGLAIKYENPREVGADRVNAVAGIHL 120
| : : |||| : | : : || : || : || : || : || : || : || : || : ;
Db 2 MLIAIEOGNNTNFATHDGASWVAQWRSAATESRTADEYVVMQLSQQSGFRDAIV 61
| : : |||| : | : : |||| : || : || : || : || : || : || : || : ;

QY 61 IVSVVPPIMPALERCEKEFKPLPVVGGIKTGLAIKYENPREVGADRVNAVAGIHL 120
| : : |||| : | : : |||| : || : || : || : || : || : || : || : ;

Db 62 IIISVQPQSIPFNLRNSRRYFNVEPLVIGENAKLGIDVRIEKPSGEAGADRLVNAIGAMV 121
| : : |||| : | : : |||| : || : || : || : || : || : || : || : ;

QY 121 YGSLPIVDGTATTVCYINEEKHYMGVITPGLIMISAEALYSRAAKLPRIETKPSS-- 178
| : : |||| : | : : |||| : || : || : || : || : |||| : | : ;

Db 122 YPGPLVIDSGTATTFDIVAADGAFCGIATPGLINSMQALHEAAKLPRATIOGPAGNR 181
| : : |||| : | : : |||| : || : || : || : || : |||| : | : ;

QY 179 VQGKTVSAMOSGILGYGVQEVGIVKRMKEEAQEPKIATGGLAKLISESNVIDYVD 238
| : : |||| : | : : |||| : || : || : || : || : |||| : | : ;

Db 182 IVGTDTVSAMOSGVFWCYISLIEGLVARIKAEERPEMTVIATGCVASLFEGATDSIDHFD 241
| : : |||| : | : : |||| : || : || : || : || : |||| : | : ;

QY 239 PFLTLKGLMYLYERN 253
| : : |||| : | : : |||| : || : || : || : || : |||| : | : ;

Db 242 SDLTIRGLLEIYRRN 256
| : : |||| : | : : |||| : || : || : || : || : |||| : | : ;

RESULT 10

Q9CD56 PRELIMINARY; PRT; 274 AA.

ID Q9CD56
AC Q9CD56;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ML0232.
GN ML0232.

OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]

SQ SEQUENCE FROM N.A.

THIS PAGE BLANK (USPTO)

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

[illegible]

```

Db      61    IISVVPPPIHMAHEMTCVRYFNIRPLIVPGIGIKTGLNLKVDNPREGTSRIVNAVAASEE 120
QY     121 YRTSLIIVDFGTTATFDYYNRRGEYCGGAIAAPGLVIISTEALFORASKLPRVDIIIRPSAII 180
Db     121 YGTPPVIVDFGTATTFCYIDESCVOGGAIPGMISTEALYNRAAKLFRVIDAESQII 180
QY     181 ARNTVNMSMGAGIYGVGVLVDEIVTRMKAEKDAPRVATGGIASLIAPESTKTIEAVEY 240
Db     181 GKSTVSSMQAGIFYGFVGCQEGIIAEMKKOSNASPVVWATGGILARMITEKSSAVDILDPF 240
QY     241 LTLGLELRILYERNR 254
Db     241 LTLKGLELLYRNK 254

RESULT 4
AF1464
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip1126)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1464
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkhat, G.; Madueno, E.; Maitournam, A.;
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weha
A:title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:2137279; PMID:11679669
A:Accession: AF1464
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin0253
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94_31c

Query Match          58.2%; Score 745; DB 2; Length 259;
Best Local Similarity 54.3%; Pred. No. 2.4e-55;
Matches 138; Conservative 54; Mismatches 62; Indels 0; Gaps 0;

QY     1 MLIVIDGVNTNVLGDIYDCERLVDRWRVSTDKARTTDGYGIILNELFRLAGLGLDQIRAV 60
Db     1 MLIVIDGVNTCTVGYKEQLKLHRWTTHRHVSDELGMTVLNFESVANLTSPSDIQGI 60
QY     61 IISVVPPPTUGVLERSLGFGMRPLIVPGIGIKTGMPITQYDNPREVGADRIYNVAVGYEK 120
Db     61 IISVVPPPIHMAHEMTCVRYFNIRPLIVPGIGIKTGLNLKVDNPREGTSRIVNAVAASEE 120
QY     121 YRTSLIIVDFGTTATFDYYNRRGEYCGGAIAAPGLVIISTEALFORASKLPRVDIIIRPSAII 180
Db     121 YGTPPVIVDFGTATTFCYIDEAGVYOGGAIPGMISTEALYNRAAKLFRVIDAESQII 180
QY     181 ARNTVNMSMGAGIYGVGVLVDEIVTRMKAEKDAPRVATGGIASLIAPESTKTIEAVEY 240
Db     181 GKSTVSSMQAGIFYGFVGCQEGIIAEMKKQSNTSPVWATGGILARMITEKSSAVDILDPF 240
QY     241 LTLGLELRILYERNR 254
Db     241 LTLKGLELLYRNK 254

RESULT 5
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C:Accession: S66100; E69740
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A>Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis ch

```

A:Reference number: S65967; MUID:96051385; PMID:7584024
A:Accession: S66100
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <G>
A:Cross-references: EMBL:D26185; NID:9467326; PIDN:BAA05305.1; PID:dl005847; PID:9467459
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.: Bron, S.; Broutlet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Faret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galled
teich, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, R.; Sekiguchi, J.; Sekowska, A.; Seron
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69740
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <KUN>
A:Cross-references: GB:299104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:ell182003;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yacB
A:Start codon: TTG
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 50.8%; Score 650; DB 2; Length 233;
Best Local Similarity 57.4%; Pred. No. 2.1e-47;
Matches 124; Conservative 38; Mismatches 54; Indels 0; Gaps 0;

QY 1 MLLVDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGLDQIRAV 60
DB 1 MLLVDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGLDQIRAV 60

QY 61 IISSVVPPLTGVLRSLGFGMRPLVVGPGIKTGMPIQYDNPREVGADRIYNAVAGYEK 120
DB 61 IISSVVPPLTGVLRSLGFGMRPLVVGPGIKTGMPIQYDNPREVGADRIYNAVAGYEK 120

QY 121 YRTSLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDIIRPSAII 180
DB 121 YRTSLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDIIRPSAII 180

QY 121 YGNPLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDIIRPSAII 180
DB 121 YGNPLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDIIRPSAII 180

QY 181 ARNTVNSMQAGIYGYGVGLVDEIVTRMKAESKDAPR 216
DB 181 KNTVNSMQAGIYGYGVGLVDEIVTRMKAESKDAPR 216

RESULT 6
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36391
R:Oliver, K.; Harris, S.D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T36391
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-265 <OLI>
A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE94.31c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 49.2%; Score 629; DB 2; Length 265;
Best Local Similarity 51.1%; Pred. No. 1.5e-45;
Matches 134; Conservative 42; Mismatches 76; Indels 10; Gaps 5;

QY 1 MLLVDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGLDQIRAV 55
DB 1 MLLVDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGLDQIRAV 55

QY 56 QIRAVISSVVPPLTGVLRSLGFGMRPLVVGPGIKTGMPIQYDNPREVGADRIYNAV 114
DB 56 QIRAVISSVVPPLTGVLRSLGFGMRPLVVGPGIKTGMPIQYDNPREVGADRIYNAV 114

QY 60 GIDGTAICATPVSVLHRELRVRYGDPVAVLVEPVGVTKTGPVITDHPKEVGADRIINA 119
DB 60 GIDGTAICATPVSVLHRELRVRYGDPVAVLVEPVGVTKTGPVITDHPKEVGADRIINA 119

QY 115 VAGYEKRTSLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDI 174
DB 115 VAGYEKRTSLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDI 174

QY 120 VAAVELYGGPAIVDFGTATFDVARSARGEYIGVIAFGIEISVALGVKQALKEIVA 179
DB 120 VAAVELYGGPAIVDFGTATFDVARSARGEYIGVIAFGIEISVALGVKQALKEIVA 179

QY 175 RPSAIIARTVNSMQAGIYGYGVGLVDEIVTRMKAESKDAP---RVIIATGGIASIAPES 231
DB 175 RPSAIIARTVNSMQAGIYGYGVGLVDEIVTRMKAESKDAP---RVIIATGGIASIAPES 231

QY 180 RPRSIVIGKNTVEAMQSGIVYGFAGQVGVNRMARLADDPDDVTVIATGGLAPMVLGES 239
DB 180 RPRSIVIGKNTVEAMQSGIVYGFAGQVGVNRMARLADDPDDVTVIATGGLAPMVLGES 239

QY 232 KTIEAVEEYLTLEGLRLIYERN 253
DB 232 KTIEAVEEYLTLEGLRLIYERN 253

QY 240 SVIDEHEPWLTLMLGLRLIYERN 261
DB 240 SVIDEHEPWLTLMLGLRLIYERN 261

RESULT 7
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: B87489
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, K.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-261 <STO>
A:Cross-references: GB:AE005673; NID:gi13423392; PIDN:AAK23391.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1935
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 43.9%; Score 561; DB 2; Length 261;
Best Local Similarity 44.7%; Pred. No. 7.7e-40;
Matches 114; Conservative 50; Mismatches 89; Indels 2; Gaps 1;

QY 1 MLLVDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGLDQIRAV 60
DB 1 MLLVDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGLDQIRAV 60

QY 61 IISSVVPPLTGVLRSLGFGMRPLVVGPGIKTGMPIQYDNPREVGADRIYNAVAGYEK 120
DB 61 IISSVVPPLTGVLRSLGFGMRPLVVGPGIKTGMPIQYDNPREVGADRIYNAVAGYEK 120

QY 121 YRTSLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDIIRPSAII 178
DB 121 YRTSLIIVDFGTATFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVRVDIIRPSAII 178

QY 122 YPGPLVIVDSGTATFDVAAAGPEGGIAPGINLSMQALHEAAKLPRIATQIPAGNR 181
DB 122 YPGPLVIVDSGTATFDVAAAGPEGGIAPGINLSMQALHEAAKLPRIATQIPAGNR 181

QY 179 IIARTVNSMQAGIYGYGVGLVDEIVTRMKAESKDAPRVIIATGGIASIAPESKTEAVE 238
DB 179 IIARTVNSMQAGIYGYGVGLVDEIVTRMKAESKDAPRVIIATGGIASIAPESKTEAVE 238

QY 182 IVGDTVNSMQSGVFWGISLIEGLVARIKAEGERPMTVIATGVASLFEAGTDSIDHFD 241
DB 182 IVGDTVNSMQSGVFWGISLIEGLVARIKAEGERPMTVIATGVASLFEAGTDSIDHFD 241

QY 239 EYLTLEGLRLIYERN 253
DB 239 EYLTLEGLRLIYERN 253

QY 242 SOLTIRGLLEIYERN 256
DB 242 SOLTIRGLLEIYERN 256

RESULT 8

Qy	1	MLLIVDVGNNTNVLGIYDG-----ERLVRDNRVSTDKARTDEYGLLINELFRLAGLGDQ	56
Db	1	MLLIVDVGNNTNVLGIYDG-----ERLVRDNRVSTDKARTDEYGLLINELFRLAGLGDQ	56
Qy	57	IRLAVIISVVVPLTGLVLERLSLGYFCMRP-LVVGPGIKTGMPIQVODNPREVGADRIYNVAV	111
Db	58	LTCSTAALSTVPSVLHREVRIMLDOYSPSPVPHVLIEPGVGTGIPLLVDNPKYVGADRIYNVCL	111
Qy	116	AGYEKRTSLIIVDFGTATTFDYVNRKGYCCGALAPGLVISTEALFORASKLPRVDIIR	177
Db	118	AAVDRFKRAIIVDFGSSICVDVSNKGEFLGATAPGVQVSSDAAAARSAAARVELAR	177
Qy	176	PSAIIARTNWSNQAGIYGYVGLVDEIVTRMAE-----SKDAP-RVIATCGGLASLIAPE	231
Db	178	PRSVGKNTVECMQAGAFEGFAGLVLDGLVGRIREDSVGSFVDHDAIVATGHTAPLLIPE	231
Qy	231	SKTIEAVEEYLLIEGRILYERNRE	255
Db	238	LHPTVDHYDQHLTLQGLRLVFERNLE	262

RESULT 10
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: E75516
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodsworth, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75516
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <WHI>
A:Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; F:
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0461
A:Map position: 1
C:Superfamily: Streptomycetes coelicolor hypothetical protein SCE94.31c

Best Local Matches	Similarity	36.8%;	Pred. NO. 1.7e-27?	Mismatches	50;	Indels	13;	Gaps	6
Qy	2	LLAVIDVGNINVLGIYDGE-RLVDRWRVSTDKARTTDEYIGILINELFRLAGLGLDQIRAV	60						
Db	6	LLAVDIGNTTVLGLADASGALTHTWRITNREMLPDDLALQLHGLFTLACAPIP--RAA	63						
Qy	61	LISSVVPPLTGLVLERLSLG--YFGMRPLVVGPGIKTCMPLOYDNPREVGADRLVNNAVAG	117						
Db	64	VLISSVAPP---VGENVALKAKRHFMIDAFSAENLPDVTVELDTPGSGADRLCN-LFG	119						
Qy	118	YEKIRFSL---LIIVDFGTAFTFDYVNRKGEYCGGAIAPGLVISTEALFQRASKLPVRVDII	174						
Db	120	AEKYLGGDLVAVVDFGFTSTNFDVYGRRRFLTGGTLATGAQVSADALFARAAKLPRTLTQ	179						
Qy	175	RFSAITARTNVSMOAGIYGVGVGLVDDELVTNRKAEKSDAPRVATGTGLASLIAPEKSTI	233						
Db	180	APETAIGKNTVHALQSLGVFGYAEVMDGLLRRIRAEPLGGEAVAVATGFGFSTVVGICQEI	233						
Qy	235	EAVEEYLTLEGLRILYERNRE	255						
Db	240	DYDEFTLTLLRGLVFWASRSE	260						

RESULT 11
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C:Accession: D71326

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998
A:Title: Complete genome sequence of *Treponema pallidum*, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: D71326

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <COL>

A:Cross-references: GB:AE001220; GB:AE000520; NID:G3322705; PIDN:AA65417.1; PID:G332271

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0431

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 29.0%; Score 371.5; DB 2; Length 273;

Best Local Similarity 34.7%; Pred. No. 7.5e-24;

Matches 87; Conservative 56; Mismatches 101; Indels 7; Gaps 4;

Qy 2 LLVIDGNTNVLGIY---DGERLVRD-WRVSTDKARTDEYGLINELFLRAGLGDOI 57

Db 1 MLLIDVGNHNVFGIOGNGRVCVRELRLAPDARKTQDEYSLIHALCERAGVGRASL 60

Qy 58 RAVIISVWPPLTGVLRLSLGYFGMRPLVVGPIGTGMPIOYDNP--REVGADRIYNAV 115

Db 61 RDAFISSVPLTKTIADVAQISGVQPVFGPWAYEHLPRPEPRAEIGTDLVANAV 120

Qy 116 AGYEKRTSLIIVDFGATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIR 175

Db 121 AAYVHFRSACVVDCGTALFTAVDGTGLIQGVAIPGLRTAVQSLHTGTQAQLPLVPLAL 180

Qy 176 PSALIIARTNVMQAGIYGVGLVDEIVTRMKESKADPRVIATGGLASLIAPESKTIE 235

Db 181 PDSVLGKDTHAVQAGVGRVTLFVIRAMIAQCCKELGCRCAAVITGSLRFSSE-VDFP 239

Qy 236 AVEEYLTLEGL 246

Db 240 PIDAQLTSLGL 250

RESULT 12

D72320

conserved hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

:Accession: D72320

:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: D72320

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-246 <ARN>

A:Cross-references: GB:AE001754; GB:AE000512; NID:94981417; PIDN:AD35964.1; PID:9498141

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0883

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.3%; Score 361.5; DB 2; Length 246;

Best Local Similarity 36.8%; Pred. No. 4.6e-23;

Matches 91; Conservative 49; Mismatches 98; Indels 9; Gaps 6;

Qy 1 MLLVIDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGDOIIRAV 60

Db 1 MLLVIDGNTNHSVFSTEDGKTRFRWLSTGVFQTEDE---LFSHLPLGLGDAMREIKGI 57

Qy 61 IISVVPPLTGVLRLSLGYFGMRPLVVGPIGTG-MPIQYDNPREVGADRIYNAVAGYE 119
Db 58 GVASVWQTQNTVIERFSQYKPHISPIWV--KAKNGCVKNVKNSEVGDARVANVAVRK 115
Qy 120 KYRSLIIVDEGTATTFDYVNRKGEYCGGATAPGLVISTEALFORASKLPRVDIIRPSAI 179
Db 116 EYKNGIIDIIMGTATTVDLV-VNGSYEGGATLPGFFMMVHSLFRGTAKLPLVEVKPADFV 174
Qy 180 IARTNVMQAGIYGVGLVDEIVTRMKESKADPRVIATGGLASLIAPESKTIEAVEE 239
Db 175 VGKDFEENIRLGCVNGSVYALEGLIGRIKEYVGDLPL-VLTGGQSKIYKDMIKH-EIFDE 232
Qy 240 YLTLEGL 246
Db 233 DLTIKGV 239

RESULT 13

F70165

conserved hypothetical protein BB0527 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999

C:Accession: F70165

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh

son, D.; Peterson, J.; Kervavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu

bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.

A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: F70165

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-262 <KLE>

A:Cross-references: GB:AE001154; GB:AE000783; NID:G2688431; PIDN:AAC66882.1; PID:G268

A:Experimental source: strain B31

Query Match 20.8%; Score 266.5; DB 2; Length 262;

Best Local Similarity 29.3%; Pred. No. 5e-15;

Matches 75; Conservative 48; Mismatches 112; Indels 21; Gaps 5;

Qy 3 LVIDGNTNVLGIYDGERLVRDWRVSTDKARTDEYGLINELFLRAGLGDOIIRAVII 62

Db 9 LIIDIGNTSIAFALFKDNQVNLFIKMTNLMRYDEVYSFFPENFDF-----NVNKFVI 62

Qy 63 SSVVPLTGVLRLSLGYFGMRPLVVGPIGTGMPIOYD---NPRE-----VGADRIYNA 114

Db 63 SSVVPLINETPKNVIFFFKIKPLFI-----GFDLNYDLTFNPKYDKFLGSDVFNAL 116

Qy 115 VAGYEKVR-TSLIIVDFGATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDI 173

Db 117 VAATENYSFENVVLDLGTACTIFAVSQDGLGILGINSGLPLNFNSLDNAYLKKFPI 176

Qy 174 IRPSAIIARTNVMQAGIYGVGLVDEIVTRMKESKADPRVIATGGLASLIAPESKT 233

Db 177 STPNLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKFNLIIITGKNADLILSLIEI 236

Qy 234 IEAVEEYLTLEGLRIL 249

Db 237 EFIFNIHLTVGVRIL 252

RESULT 14

140327

baf protein - Bordetella pertussis

C:Species: Bordetella pertussis

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999

C:Accession: I40327; S70569

R:DeShazer, D.; Wood, G.E.; Friedman, R.L.

J. Bacteriol. 177, 3801-3807, 1995

A:Title: Identification of a Bordetella pertussis regulatory factor required for tran

A:Reference number: I40327; MUID:95325323; PMID:7601846

A:Accession: I40327

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-267 <RES>

A:Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229

R:Allen, A.; Maskell, D.

Mol. Microbiol. 19, 37-52, 1996

A:Title: The identification, cloning and mutagenesis of a genetic locus required for ltr

A:Reference number: S70669; MUID:96419162; PMID:8821935

A:Accession: S70669

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 239-267 <ALL>

A:Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

C:Genetics:

A:Gene: baf

Query Match 11.6%; Score 148; DB 2; Length 267;
Best Local Similarity 23.1%; Pred. No. 4.9e-05;
Matches 66; Conservative 46; Mismatches 110; Indels 64; Gaps 12;

1 MLIVDVGNTNIVLGIYDGE--RLVRD-----WRYSTDKARTTDEYGL 42

1 MIIIDSGNSRUKGVWFDPAQAREPAPAFDNDLDALGRW-LATLPRRQALGV- 58

43 INELFRLAGLIGLQIRAVIISVVPLTGVLRSLGFGMRPLVVGPGIKTGMPIQYDN 102

59 -----NVAGLARGEATAVIRA-----GCCDIRWLRAQPLAM--GLRNG-----YRN 98

103 PREVGAADR---IVNAVAGYKRTSLIIVDFGTATTFDYVNRKGYCGGAIAPGLVISTE 159

99 PDQLGADRACWAGVGLRQPSVHPPLLVASFGTATTTDTPDNPVPPGGLILPGPAMRG 158

160 ALFQRASKLPVDDIIRPSAIARNTVNSMQA---GIYGVVGLVDEIVTRMKAESKDAPR 216

159 ALAYGTAHLPLAD-----GLVADYPIDTHQAIASGAAQAAGVIRQWLAGRORYGQAPE 213

217 VIATGGLASLIAPESKTIEAV-----BEYL---TLEGRLIL 249

214 IYVAGGWPVROEAERLLAVTGAAGFATPQPTIYLDSPVLDGLAAL 259

RESULT 15

H83111

hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: H83111

Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bu

man, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: H83111

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-248 <STO>

A:Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AG07667.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4279

Query Match 11.0%; Score 141; DB 2; Length 248;
Best Local Similarity 25.1%; Pred. No. 0.00017;
Matches 67; Conservative 38; Mismatches 110; Indels 52; Gaps 10;

1 MLIVDVGNTNIVLGIYDGERLVRWRVSTDKART-----TDEYGLINELFRLAGLID 55

1 MILEDCGNSLI-----KWRVIEGAARSVAGGLAESDDALVEQLTSQQLPVR 48

56 QRRAVIIS--VVPPLTGVLRSLGFGMRPLVVGPGIK--TGMPIQYDNPREVGADRIV 112

Db 49 ACRLSVSRSEQETSQVLARLEQL-----FPVSALVASSGKQLAGVRNGYLDYQRLGLDRWL 104

QY 113 NAVAGYKRTSLIIVDFGTATTFDYVNRKGYCGGAIAPGLVISTEALFQRASKLPVDD 172

Db 105 ALVAAHHLAKKACLVLDLGTAVTSDLVAAADGVHLGGYICPGMTLMRSQURTHTRR----- 159

QY 173 IIRPSAIIARNTVNSMQA-----IYGVVGLVDEIVTRMKAESKD--AP--RVATATGG 222

Db 160 -TRYDDAEARRALASLOPQOATAEAVERGCLLMRLGFRVREQYAMACELLGPDCEIFLTGG 218

QY 223 LASLIAPESKTIIEAVEEYLTLEGRLIL 249

Db 219 DAELVRDE-----LAGARIM 233

Search completed: June 24, 2003, 22:03:03

Job time : 10.3009 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.43516 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-7

Perfect score: 1279

Sequence: 1 MLLVIVGNTNIVLGIYDGE.....AVEEYLTLEGLIRLYERNRE 255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	650	50.8	233	1 YACB_BACSU	P37564 bacillus su
2	148	11.6	267	1 BAF_BORPE	Q45338 bordetella
3	98.5	7.7	256	1 HI61_PSEAE	Q9hu44 pseudomonas
4	97.5	7.6	375	1 BUK2_THEMA	Q9x278 thermotoga
5	97	7.6	391	1 RRPP_MUMPM	P16595 mumps virus
6	94	7.3	610	1 FTSI_HAEIN	P45059 haemophilus
7	94	7.3	758	1 LEU2_SCHPO	Q4289 schizosacch
8	93	7.3	872	1 DNAB_SYNY3	Q5418 synechocyst
9	92	7.2	297	1 PUR7_MYCTU	Q59566 mycobacteri
10	89.5	7.0	390	1 RRPP_MUMPI	P19717 mumps virus
11	89.5	7.0	587	1 DGOA_ECOLI	P14558 e dgoa prot
12	88.5	6.9	947	1 PMA2_YEAST	P19657 saccharomyc
13	87.5	6.8	469	1 LEU2_RHIME	Q2176 rhizobium m
14	87	6.8	391	1 RRPP_MUMPE	P16072 mumps virus
15	86	6.7	296	1 HSI1_YARLI	Q99145 yarrowia li
16	86	6.7	437	1 SGCC_ECOLI	P39365 escherichia
17	85.5	6.7	336	1 PYRD_SALTY	P25468 salmonella
18	84.5	6.6	555	1 VGLF_PILHC	P12605 human parai
19	84.5	6.6	750	1 LEU2_RHINI	P35811 rhizopus ni
20	84.5	6.6	840	1 MUTS_LACLA	Q9cdk9 lactococcus
21	83.5	6.5	336	1 PYRD_ECOLI	P05021 escherichia
22	83.5	6.5	403	1 DPP_METUA	Q38323 methanococc
23	83.5	6.5	416	1 PGK_HORSE	P00559 equus cabal
24	83	6.5	356	1 BUK_CLOPE	Q2953 clostridium
25	83	6.5	419	1 PGK_TREPA	Q83549 treponema p
26	82.5	6.5	469	1 LEU2_BRUME	Q9yjc9 bruceella me
27	82.5	6.5	486	1 GLGA_THEMA	Q9wzz7 thermotoga
28	82.5	6.5	553	1 VGLF_NDVI	P14623 newcastle d
29	82	6.4	253	1 HIS6_PYRAE	Q8zy16 pyrobaculum
30	82	6.4	412	1 IDHC_YEAST	P41939 saccharomyc
31	82	6.4	568	1 CH60_CABEL	P50140 caenorhabdi
32	82	6.4	1344	1 XDH_DROSU	P91711 drosophila
33	81.5	6.4	416	1 PGK_CRIGR	P50310 cricetus

34	81.5	6.4	470	1 RHBA_RHIME	Q9z3r2 r diaminobu
35	81.5	6.4	621	1 GIDA_HELPJ	Q9zml9 helicobacte
36	81.5	6.4	713	1 RIR3_ECOLI	P39452 escherichia
37	81.5	6.4	849	1 PHSG_SYNY3	P73511 synechocyst
38	81	6.3	398	1 KBL_SALTY	P74119 salmonella
39	81	6.3	664	1 NTP1_ENTHR	P43439 enterococu
40	81	6.3	778	1 FTSK_COXBU	P39920 coxiella bu
41	80.5	6.3	286	1 PUR7_PASBU	P57878 pasteurella
42	80.5	6.3	398	1 KBL_ECOLI	P07912 escherichia
43	80.5	6.3	458	1 RADA_BACSU	P37572 bacillus su
44	80.5	6.3	469	1 LEU2_AGRF5	Q8uby9 agrobacteri
45	80.5	6.3	540	1 GTR9_HUMAN	Q9nrm0 homo sapien

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gham S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche R., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler F., Vassarotti A.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zunstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

```
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
DR EMBL; D26185; BAA05305.1; -
DR EMBL; Z99104; CAB11846.1; -
DR Subtilisin; BG10133; YacB.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26217 MW; AAE56E732C15DF44 CRC64;

Query Match 50.8%; Score 650; DB 1; Length 233;
Best Local Similarity 57.4%; Pred. NO. 1.9e-47;
Matches 124; Conservative 38; Mismatches 54; Indels 0; Gaps 0;

QY 1 MLLIVDVGNTNVLGIDYGERLVRVSTDKARTTDEYGLINELERLAGLGLDQIRAV 60
DB 1 MLLIVDVGNTNVLGIDYGERLVRVSTDKARTTDEYGLINELERLAGLGLDQIRAV 60
QY 61 IISVVPPITGVLERLSLGYFGMRPLVGVGPKTGMPIQYDNPREVGDRIYNVAGYEK 120
DB 61 IISVVPPITGVLERLSLGYFGMRPLVGVGPKTGMPIQYDNPREVGDRIYNVAGYEK 120
QY 121 YFTSLIVDFGATTFDYNVRKGYCGGAIAPGLVISTEALFORAKSLPRVDIIRPSAI 180
DB 121 YFTSLIVDFGATTFDYNVRKGYCGGAIAPGLVISTEALFORAKSLPRVDIIRPSAI 180
QY 121 YGNPLVDFGATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
DB 121 YGNPLVDFGATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
QY 181 ARNTVNSMAGLYGVGLVDELIVTRMKAESKDAPR 216
DB 181 GRNTVSAMSGILFVGIVGQVEGIVRRMKWQAKQDPR 216

RESULT 2
BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
Bordetella pertussis.
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP504;
RX MEDLINE-95325323; PubMed-7601846;
RA Deshaizer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
for transcription of the pertussis toxin operon in Escherichia
coli."
RT J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN-BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a bira homolog in Bordetella pertussis."
RN [3]
RP Submitted (JUL-1997) to the EMBL/GenBank/DDBB databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN-BP536;
RX MEDLINE-96419162; PubMed-8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus
```

```
required for lipopolysaccharide biosynthesis in Bordetella
pertussis."
Mol. Microbiol. 19:37-52(1996).
-1- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
RNA POLYMERASE.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U12020; AAY75361.1; -
DR EMBL; AF016461; AAC68834.1; -
DR EMBL; X90711; CAA62242.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match 11.6%; Score 148; DB 1; Length 267;
Best Local Similarity 23.1%; Pred. NO. 2.4e-05;
Matches 66; Conservative 46; Mismatches 110; Indels 64; Gaps 12;

QY 1 MLLIVDVGNTNVLGIDYGE--RLVRD-----WRVSTDKARTTDEYGL 42
DB 1 MLLIVDVGNTNVLGIDYGE--RLVRD-----WRVSTDKARTTDEYGL 42
QY 43 INELERLAGLGLDQIRAVIISVVPPITGVLERLSLGYFGMRPLVGVGPKTGMPIQYDN 102
DB 43 INELERLAGLGLDQIRAVIISVVPPITGVLERLSLGYFGMRPLVGVGPKTGMPIQYDN 102
QY 59 -----NVAGLARGEIAATLRA-----GCCDIRLWLRQAOLAM--GLRNG----YRN 98
DB 59 -----NVAGLARGEIAATLRA-----GCCDIRLWLRQAOLAM--GLRNG----YRN 98
QY 103 PREVGADR---IVNAVAGYEKRTSLIIVDFGTATTFDYNVRKGYCGGAIAPGLVISTE 159
DB 103 PREVGADR---IVNAVAGYEKRTSLIIVDFGTATTFDYNVRKGYCGGAIAPGLVISTE 159
QY 99 PDQAGADRWACWGVGLARQPSVHPPLLVASFGTATTLDTIGPDNVFPGLLPGPAMMRG 158
DB 99 PDQAGADRWACWGVGLARQPSVHPPLLVASFGTATTLDTIGPDNVFPGLLPGPAMMRG 158
QY 160 ALFORASKLPYVDIIRPSAIIRNTVNSMQA---GIYGVGLVDELIVTRMKAESKDAPR 216
DB 160 ALFORASKLPYVDIIRPSAIIRNTVNSMQA---GIYGVGLVDELIVTRMKAESKDAPR 216
QY 159 ALAYGTAHLPLAD-----GLVADYDIDTHQATASGIAAQAQAIYRWLAGRQRYGQAP 213
DB 159 ALAYGTAHLPLAD-----GLVADYDIDTHQATASGIAAQAQAIYRWLAGRQRYGQAP 213
QY 217 VIATGGLASLIAPESKTIEAV-----BEYL---TLEGLRIL 249
DB 217 VIATGGLASLIAPESKTIEAV-----BEYL---TLEGLRIL 249
QY 214 IVVAGGWEPEVQEAERLAVTGAAGFATPQPTLYLDSPLVDGLAAL 259
DB 214 IVVAGGWEPEVQEAERLAVTGAAGFATPQPTLYLDSPLVDGLAAL 259

RESULT 3
HIG1_PSEAE STANDARD; PRT; 256 AA.
ID HIG1_PSEAE
AC Q9HU44;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Imidazole glycerol phosphate synthase subunit hisF1 (EC 4.1.3.-) (IGP
synthase cyclase subunit) (IGP synthase subunit hisF1) (ImGP synthase
subunit hisF1) (IGPS subunit hisF1).
GN HISF1 OR PA5140.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Miziouchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
```


RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RL opportunistic pathogen.";
 CC Nature 406:959-964(2000).

CC -I- FUNCTION: IGS catalyzes the conversion of PRPAR and glutamine to
 CC cyclization and glutamate. The hsf subunit catalyzes the
 CC cyclization activity that produces IGP and AICAR from PRPAR using
 CC the ammonia provided by the hsf subunit (By similarity).

CC -I- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
 CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
 CC carboxamide + L-glutamine -> imidazole-glycerol phosphate + 5-
 CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.

CC -I- PATHWAY: Histidine biosynthesis; fifth step.

CC -I- SUBUNIT: Heterodimer of hsf and hsf (By similarity).

CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -I- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE004927; AAC08525.1; -

DR InterPro; IPR003009; FHN_enzyme.

DR Pfam; PF00977; His_biosynth; 1.

DR TIGRfams; TIGR00735; hsf; 1.

KW Histidine biosynthesis; Lyase; Complete proteome.

FT ACT_SITE 12 12 POTENTIAL.

FT ACT_SITE 131 131 POTENTIAL.

SQ SEQUENCE 256 AA; 27131 MW; 6486A76CD308FAA7 CRC64;

Query Match 7.7%; Score 98.5; DB 1; Length 256;

Best Local Similarity 24.6%; Pred. No. 0.32;

Matches 62; Conservative 36; Mismatches 75; Indels 79; Gaps 14;

QY 5 IDVGNITVLGI-YDGERLVRDVRVSTDKARTDEYGLINELFRLAGLGLDQIRAVIIS 63

DB 11 LDVDNGRVKGVKEN---INDAGDPVEIARYDEQ-----GADEITFDIT 54

QY 64 SVVPLPGLV---ERLSLGYGMRPLVVGPGIKTGMPIQYDNPREGADRIVNAVGYEK 120

DB 55 ASVDGRDITLTVRMASQVF--IPLAVGGVRSVQDIR--NLLNAGADK----- 100

QY 121 YRTSLIIVDFCTATF-----DIVNRKGEYCGGAIAPGLVISTEALFQRA-SKLPRVDI 173

DB 101 -----VSINTAAVFNPFEVGEAADRFSGQ-----IVVAIDAKVSPAGAPRWEI 146

QY 174 I-----RPSAIARTNVSM-----QAGIYYGYVGLVDEIVTRMKAESKDA 214

DB 147 FTHGGRPTGLDAVLWAKMEDLGLAGEILLTSMDDGVKSGY----DLGVTRAISEAVNV 202

QY 215 PRVIATGGLASL 226

DB 203 P-VIASGGVGNL 213

RESULT 4

BUK2_THEME

ID BUK2_THEME STANDARD; PRT; 375 AA.

AC Q9X278;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable butyrate kinase 2 (EC 2.7.2.7) (BK 2) (Branched-chain

DE carboxylic acid kinase 2).

GN BUK2 OR TM1756.

OS Thermotoga maritima.

OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;

OC Thermotogaceae; Thermotoga.

OX NCBI_TaxID=2336;

RN SEQUENCE FROM N.A.
 RC STRAIN=MS8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RL genome sequence of *Thermotoga maritima*."

CC -I- CATALYTIC ACTIVITY: ATP + 2-butanolate -> ADP + butanoyl phosphate.

CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -I- SIMILARITY: BELONGS TO THE ACETOKINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE001814; AAD36821.1; -

DR TIGR; TM1756; -

DR InterPro; IPR000890; Acetate_kin.

DR Pfam; PF00871; Acetate_kinase; 1.

DR PRINTS; PR00471; ACETATEKINASE.

DR PROSITE; PS01075; ACETATE_KINASE_1; 1.

DR PROSITE; PS01076; ACETATE_KINASE_2; 1.

KW Transferase; Kinase; Complete proteome.

SQ SEQUENCE 375 AA; 42022 MW; E42C4887071D008D CRC64;

Query Match 7.6%; Score 97.5; DB 1; Length 375;

Best Local Similarity 21.3%; Pred. No. 0.62;

Matches 75; Conservative 42; Mismatches 122; Indels 113; Gaps 15;

QY 2 LLVIDVGNITVLGIYDGERLVRDVRVSTDKARTTDEYG---ILINELFR----- 48

DB 4 ILTINPGSTKLSIFEDERMVKQNFSS---HSPDELGRFQKLDQLEFREKIAQRFVE 59

QY 49 LAGLGLDQIRAVII-SSVVPPLTG-----VLERLSLGYGMRPLVVGPGI-----K 93

DB 60 ETGYSLSFSFAVSRRGGLDPIPGGVLDGLMIKTASGRNGEASNLGAIARFSSE 119

QY 94 TGMPIQYDNPREGADRIVNAVGYEKYRTSLIIVDFGTATT-----FDYVNRK 142

DB 120 TGVPAVYVDPVVVDEMEDVARVSGHPNTQRKSIFHALNQKTVAKVARMNKRVEEMNLV 179

QY 143 GEYCGGATA---PGLVISTEALF-----QRASKLPRVDIIRPSAIARTNVSMQ 189

DB 180 VAHMGGSISIAAHRKGRVIDVNNALDGDGPFPTSPERSGLTQLV-DLCFSKFTYEEMK 238

QY 190 A-----GIYYGYVGLVD--EIVTRMKAESKAPRV----- 217

DB 239 KRIVNGGLVAYLGTSDAREVVRRLKQGDWAKRYRMAVYQIAKWIGMAVLKGEVDF 298

QY 218 -IATGGLAS-----LIAPEKTIIEAVEEYLFLEGLRIL 249

DB 299 IVLTGGLAHEKEFLVPWTKRVSTFAPVLVFPGSNE-----EKALALSALRVL 346

RESULT 5

RRPP_MUMPM

ID RRPP_MUMPM STANDARD; PRT; 391 AA.

AC P16595;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE RNA polymerase alpha subunit (EC 2.7.7.48) (Nucleocapsid

DE phosphoprotein).

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D64003; BAA10516.1; -
DR HSSP; P03005; 1JWE.
DR InterPro; IPR001198; DnaB_helicase.
DR InterPro; IPR003586; Hedgehog_hintC.
DR InterPro; IPR003587; Hedgehog_hintN.
DR InterPro; IPR002203; Intein.
DR InterPro; IPR004042; Intein_endonuc.
DR Pfam; PF00772; DnaB; 1.
DR PRINTS; P00379; INTEIN.
DR ProDom; PD000229; RecA; 1.
DR ProDom; PD002663; DnaB_helicase; 1.
DR SMART; SM00305; HintC; 1.
DR SMART; SM00306; HintN; 1.
DR TIGRFAMS; TIGR00665; DnaB; 1.
DR PROSITE; P50818; INTEIN_C_TER; 1.
DR PROSITE; P50819; INTEIN_ENDONUCLEASE; 1.
DR PROSITE; P50817; INTEIN_N_TER; 1.
KW DNA replication; DNA-binding; ATP-binding; Hydrolase; Helicase;
KW Primosome; Autocatalytic cleavage; Protein splicing;
KW Complete proteome.
FT CHAIN 1 380 REPLICATIVE DNA HELICASE, 1ST PART.
FT CHAIN 381 809 SSP DNAB INTEIN.
FT CHAIN 810 872 REPLICATIVE DNA HELICASE, 2ND PART.
FT NP_BIND 207 214 ATP (POTENTIAL).
SQ SEQUENCE 872 AA; 97813 MW; AC97823A1CB85109 CRC64;

Query Match 7.3%; Score 93; DB 1; Length 872;
Best Local Similarity 22.1%; Pred. No. 4.2;
Matches 50; Conservative 40; Mismatches 100; Indels 36; Gaps 10;

QY 57 IRAVIVSVPLTGVLEKSLGYG---MRPLVWPGTKTGMPQYDNPREGADRI-- 111
DB 246 LRLVASESLDS-----NRLRTGHFSQAEFEPLTAAMGTLSSLIYIDDTASISVTQMS 300
QY 112 -NNAVAGYKRYTSLIIVDF-----GTATTEDYVNRKGEYCGAI---APGLVIS--T 158
DB 301 QVRRQSQEGPLGMWLDIYQLMEGSDNRVQELSKTRSLKGLAREINAPVALSOLS 360
QY 159 EALFORASRLPRVDIIRSAIARTVNSM-QAGYIVGYVGLVDE-----IVTRMKA 209
DB 361 RAVESRTNRPMWDLRESCIGSDSLISLSTGKRVSKLDELDEKDFEIVAINETMNL 420
210 ESKDAPRVATGGLASLIAPF--SKTIEAV--EEYLTLEGLILYE 251
421 ESAKVSREVCTGKLLVYILKTRIGRTIKATANHRELTIDGWKRLDE 466

RESULT 9
ID PUR7_MYCTU STANDARD; PRT; 297 AA.
AC Q59566; P77904;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)
DE (SAICAR synthetase).
GN PURC OR RV0780 OR MT0804 OR MTCY369.24.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_Taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-H37RV;
RX MEDLINE=96425868; PubMed=8828210;
RA Jackson M., Berthet F.-X., Otal I., Rauzier J., Martin C.,
RA Gicquel B., Guilhot C.;

RT "The Mycobacterium tuberculosis purine biosynthetic pathway:
RT isolation and characterization of the purC and purL genes.";
RL Microbiology 142:2439-2447(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean S., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + 5-amino-1-(5-phospho-D-
CC ribosyl)imidazole-4-carboxylate + L-aspartate = ADP + phosphate +
CC (S)-2-[5-amino-1-(5-phospho-D-ribosyl)imidazole-4-
CC carboxamido]succinate.
CC -!- PATHWAY: De novo purine biosynthesis; seventh step.
CC -!- SIMILARITY: BELONGS TO THE SAICAR SYNTHETASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U34957; AAB41456.1; -
DR EMBL; 280226; CAB02370.1; -
DR EMBL; AE006971; AAK45046.1; -
DR HSSP; P27616; 1A48.
DR TIGR; MT0804; -
DR Tuberculist; RV0780; -
DR InterPro; IPR001636; SAICAR_synt.
DR Pfam; PF01259; SAICAR_synt; 1.
DR ProDom; PD003043; SAICAR_synt; 1.
DR TIGRFAMS; TIGR00081; purC; 1.
DR PROSITE; PS01057; SAICAR_SYNTHETASE_1; 1.
DR PROSITE; PS01058; SAICAR_SYNTHETASE_2; 1.
KW Purine biosynthesis; Ligase; Complete proteome.
FT CONFLICT 124 124 A -> R (IN REF. 1).
FT CONFLICT 166 166 A -> P (IN REF. 1).
SQ SEQUENCE 297 AA; 32930 MW; 200D1EA37DF98637 CRC64;

Query Match 7.2%; Score 92; DB 1; Length 297;
Best Local Similarity 24.7%; Pred. No. 1.4;
Matches 56; Conservative 33; Mismatches 100; Indels 38; Gaps 12;

QY 34 RTTDEYGLINELFRLAGLGLQIRA--VLISSVPLTGVLEKSLGYGMRPLVWPG 91
DB 21 RVDDEHLIV-----ASDRISAYDITPDGRVLTAMSAFFG---LVDAFN 68
QY 92 IKTGMPQYDNP---EVCADRVNAVAGYKRYTSLIIVDFGTAT-TFDYVNRKGEYCG 147
DB 69 HLAGPP---DDPRIDPEVLGRALV--VRLEMLPVECVARGYLTSGLLDY-QATGKVCG 122

Db 255 DYLLIGDPSRINDLMQVWYRAGFYRGPGPILMSAIGDQALMDIKGVNAPVQWLMGGL 314
 QY 74 ERLSLGVE---GMRPLVVGPGIKTKMPYQY-----DNPREVGADRVNAY 115
 Db 315 VRDKIKAYSWGGDRPADVDIGIKTLUREIGDFTFKLNGCGEELGLDINSRAV--DAAVNTV 372
 QY 116 AGYEKTYSLTIIVDTATFDYVNRKGYCGGAIAP-----GLVISTEALF 162
 Db 373 AQIRE-----AFGNQIEFGL-----DFHGRVSAPMAKVLKLEPYRPLFIEBPVLA 419
 QY 163 QRASKLPVRVDIIRPSAIIRNTVNSMQAGIYGYVGLVDEIVTRMKAESKDAPRVATGG 222
 Db 420 EQAEYYPK-----LAAQTHIPLAAG-----ERMFSTRF-----DFKRVLEAGG 456
 QY 223 LASLAPESKTIEAVEYLTLLEGLRLYE 251
 Db 457 I-SILOPDLSHAGGITECKYKIAMAEAYD 484

RESULT 12

MA2_YEAST STANDARD; PRT; 947 AA.
 AC P19657;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Plasma membrane ATPase 2 (EC 3.6.3.6) (Proton pump 2).
 GN PMA2 OR YPL036W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89066768; PubMed=2904437;
 RA Schlesser A., Ulaszewski S., Ghislain M., Goffeau A.;
 RT "A second transport ATPase gene in *Saccharomyces cerevisiae*.";
 RL J. Biol. Chem. 263:19480-19487(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
 RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Dilius H., Dipaolo T., Dubois E., Duisterhoef A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messinguy F., Mewes H.-W., Mirtipati S., Moestl D.,
 RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
 RA Scherrens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
 RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
 RA Walsh S.V., Wambutt R., Wang Y., Wedler H., Winnett E.,
 RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI.";
 RL Nature 387:103-105(1997).
 CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
 CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
 CC ACTIVE TRANSPORT OF NUTRIENTS BY H⁺-SYMPORT. THE RESULTING
 CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALIZATION MAY MEDATE
 CC GROWTH RESPONSES
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
 CC H(+)(Out).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- MISCELLANEOUS: THERE ARE TWO PLASMA MEMBRANE ATPASES IN YEAST.
 CC THIS IS THE MINOR ISOFORM.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES). SUBFAMILY IITA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL: J04421; AAA83387.1; -;
 DR EMBL: U44030; AAB8184.1; -;
 DR FIR: A32023; PABY2P.
 DR SGD: S0005957; PMA2.
 DR InterPro: IPR001757; ATPase_E1-E2.
 DR InterPro: IPR004014; Cation ATPase.
 DR InterPro: IPR000695; H ATPase.
 DR InterPro: IPR001454; Hlgase/hydrolase.
 DR Pfam: PF00122; E1-E2_ATPase; 1.
 DR Pfam: PF00690; Cation_ATPase_N; 1.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PRINTS: PR00119; CATATPASE.
 DR PRINTS: PR00120; HATPASE.
 DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 KW Hydrolase; Hydrogen ion transport; Transmembrane; Phosphorylation;
 KW Magnesium; ATP-binding; Multigene family.
 FT DOMAIN 1 144
 FT TRANSMEM 145 165
 FT DOMAIN 166 169
 FT TRANSMEM 170 189
 FT DOMAIN 190 320
 FT TRANSMEM 321 342
 FT DOMAIN 343 353
 FT TRANSMEM 354 376
 FT DOMAIN 377 748
 FT TRANSMEM 749 767
 FT DOMAIN 768 783
 FT TRANSMEM 784 803
 FT DOMAIN 804 853
 FT TRANSMEM 854 874
 FT DOMAIN 875 886
 FT TRANSMEM 887 903
 FT DOMAIN 904 947
 FT MOD_RES 407 407
 FT METAL 667 667
 FT METAL 667 667
 FT DOMAIN 33 47
 FT DOMAIN 62 91
 FT CONFLICT 944 944 E -> D (IN REF. 1).
 SQ SEQUENCE 947 AA; 102171 MW; 24CF0D3EAFDIE9D5 CRC64;

Query Match 6.9%; Score 88.5; DB 1; Length 947;
 Best Local Similarity 23.4%; Pred. No. 11;
 Matches 65; Conservative 50; Mismatches 72; Indels 91; Gaps 18;

QY 15 GYDGERLVDRWVSTDKARTTDEYGLINELF-RIAGLGDOI-----RAVISSV---V 66
 Db 96 GYHAGVYVPEKDLSTDP-----YGLTSDEVARRRKKYGLNQMAEENSLIVKLFMFV 150
 QY 67 PBLTGVLRLSLGYGCMPLVVGPGIKTKMPYQYDNPREVG---ADRIYNVAGY-EKYR 122
 Db 151 GPIQFVMEAAAILAAGLSDW-----DVGVCALLLNASVGFQIEFQ 193
 QY 123 TSLIIVDF-----GTATFDYVNRKGYC-----GGAIAP--GLVISTE 159
 Db 194 AGSIVDELKTKLTATAT---VIRGQLIEIPANVWPGVGEIQLQESGTIAPADGRIVTD 249
 QY 160 ALFQRASKLPVRVDIIRPSAIIRNTVNSMQAGIYGYVGLVDEIVTRMKAESKDAPRV1- 218
 Db 250 CFLQ-----IDQSAI-----TGESLAEKHYG-----DEVFSSSTVKTGEPFMYVT 290
 QY 219 ATG-----GLASLAPESKTIIEAVEYLT--LEGLRIL 249
 Db 291 ATGDNFTVGRAAALVGQA-----SGVEGHFTFVLNGIGII 325

```

Matches 38; Conservative 29; Mismatches 67; Indels 29; Gaps 6;
95 GMPQYDNPREVADRIYNVAVGYEKYRTSLIIVDFGTATTDFYVNRKGEYCGGAIAPGL 154
   : : : : | | | | : : : : : : : : : : : : : : : : : : : :
97 GVEYSENDKQGIHVIVPEQEGFTLPGMTIVCGDSHTST-----HGAF--GALAHGI 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

155 -----VISTEALFORASK--LPRVDIIRPSAIIARTVTVNSMQAGIYYGYVGLVDEIV- 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
148 GTSEYEHVLAQTLLQKKAKNMLVRVDQGLPGVTAKDILAILGEI--GTAGGTCHVIE 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
205 -----TRKAESKDAPRVATGGLASLAPESKTEAVEE 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
206 PAGEAIRLSMEGRVTCVNTIEGARAGLIAPDETTPEYIKD 248
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
RRPP_MUMPE STANDARD; PRT; 391 AA.
P16072:
01-APR-1990 (Rel. 14, Created)
01-APR-1990 (Rel. 14, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
RNA polymerase alpha subunit (EC 2.7.7.48) (Nucleocapsid
phosphoprotein).
P.
Mumps virus (strain Enders).
Viruses; ssRNA negative-strand viruses; Mononegavirales;
Paramyxoviridae; Paramyxovirinae; Rubulavirus.
NCBI_TaxID=111167;
[1]
SEQUENCE FROM N.A.
MEDLINE=86299965; PubMed=3404121;
Takeuchi K., Hishiyama M., Yamada A., Suglura A.;
"Molecular cloning and sequence analysis of the mumps virus gene
encoding the P protein: mumps virus P gene is monocistronic.";
J. Gen. Virol. 69:2043-2049(1988).
-| POLYMERASE. THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
POLYMERASE. IT MAY FUNCTION IN TEMPLATE BINDING..
-| CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
{RNA}{N}.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D00351; BAA00259.1; -
PIR; J00067; RNZED.
InterPro; IPR004897; Paramyx_P_V.
Pfam; PF03210; Paramyx_P_V; 1.
Transferrase; RNA-directed RNA polymerase; Nucleocapsid;
Phosphorylation.
SEQUENCE 391 AA; 41650 MW; 2802A11A5542B7A5 CRC64;

Query Match 6.8%; Score 87; DB 1; Length 391;
Best Local Similarity 21.0%; Pred. No. 5;
Matches 59; Conservative 41; Mismatches 91; Indels 90; Gaps 13;

37 DEYGLINELFLAG-----LGDDQI-RAVIISSVVPPLTGLVERLSIGV----- 80
||| ||| : : : : : : : : : : : : : : : : : : : : : :
8 DETGLDITGMNVANHFSAPIQGTNSLSKASIIIPGAVPLIGNPEQRNIQHTASHQGS 67
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
81 -----FCMRPLVVGPP-----GIKTGMPITQYDNP----- 103
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
68 KSKGRGSGVRSIIIVPPSEANGGTOIPEPLFAQTQGGIVTV---YQDPTIQTGSYRS 124
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

104 ---REVGADRIYNVAVGYEKYRTSLIIVDF-----GTATTFDYVNRKGEYCGGAIAPGL- 154
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 VELTKIGKRMINREV--EKPRISPTVTEFRGGGAAAGQOTIOEEGIDGNGSAGASRSE 182
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

Query Match 6.7%; Score 86; DB 1; Length 296;

Search completed: June 24, 2003, 21:51:12
Job time : 6.43516 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 20.7979 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-7
Perfect score: 1279
Sequence: 1 MLLVIDVGNNTNVLGIYDGE.....AVEEYLTLEGLRILYERNRE 255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
tal number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	825	64.5	255	16	Q8R7M2	Q8R7M2 thermoanaer
2	818	64.0	258	2	Q9F985	Q9F985 bacillus st
3	771	60.3	254	16	Q9KGH5	Q9KGH5 bacillus ha
4	752	58.8	259	16	Q8XHL5	Q8XHL5 clostridium
5	752	58.8	273	16	Q97EB4	Q97EB4 clostridium
6	749	58.6	259	16	Q8VAC5	Q8VAC5 listeria mo
7	745	58.2	259	16	Q92F54	Q92F54 listeria in
8	629	49.2	265	16	Q9X8N6	Q9X8N6 streptomyce
9	561	43.9	261	16	Q9A6Z1	Q9A6Z1 caulobacter
10	522.5	40.9	274	16	Q9CD56	Q9CD56 mycobacteri
11	509.5	39.8	272	16	Q06282	Q06282 mycobacteri
12	423	33.1	256	16	Q8RFE4	Q8RFE4 fusobacteri
13	414.5	32.4	262	16	Q9RX54	Q9RX54 deinococcus
14	371.5	29.0	273	16	Q83446	Q83446 treponema p
15	361.5	28.3	246	16	Q9WZY5	Q9WZY5 thermotoga
16	338.5	26.5	212	2	Q32514	Q32514 desulfovibr

17	266.5	20.8	262	16	O51477	O51477 borrelia bu
18	210	16.4	295	16	Q8Y2M4	Q8Y2M4 ralstonia s
19	141	11.0	248	16	Q9HWC1	Q9HWC1 pseudomonas
20	141	11.0	592	16	Q9JW17	Q9JW17 neisseria m
21	140	10.9	276	16	Q8YQD7	Q8YQD7 anabaena sp
22	139.5	10.9	229	16	O67753	O67753 aquifex aeo
23	139.5	10.9	592	16	Q9JXF1	Q9JXF1 neisseria m
24	128	10.0	242	16	Q9PC14	Q9PC14 xylella fas
25	126	9.9	56	2	P94305	P94305 bacillus ps
26	126	9.9	257	16	P74045	P74045 synechocyst
27	114.5	9.0	223	16	Q9ZKY6	Q9ZKY6 helicobacte
28	110	8.6	597	17	O30225	O30225 archaeoglob
29	100.5	7.9	223	16	O25533	O25533 helicobacte
30	100	7.8	391	12	Q83618	Q83618 mumps virus
31	97	7.6	391	12	Q910S3	Q910S3 mumps virus
32	97	7.6	441	5	Q9VDL4	Q9VDL4 drosophila
33	96.5	7.5	464	16	Q928M6	Q928M6 listeria mo
34	95	7.4	370	16	Q9KVS3	Q9KVS3 vibrio chol
35	95	7.4	391	12	Q9DQ43	Q9DQ43 mumps virus
36	95	7.4	391	12	Q8QY72	Q8QY72 mumps virus
37	93	7.3	391	12	Q9J4L6	Q9J4L6 mumps virus
38	92.5	7.2	724	16	Q9ZCW1	Q9ZCW1 rickettsia
39	92	7.2	794	16	Q9A4G0	Q9A4G0 caulobacter
40	91.5	7.2	320	2	Q53626	Q53626 streptomyce
41	91.5	7.2	802	16	Q98CQ1	Q98CQ1 rhizobium 1
42	91	7.1	332	17	O8T761	O8T761 methanopyru
43	91	7.1	558	16	O8RAT0	O8RAT0 thermoanaer
44	90.5	7.1	341	16	Q98CD7	Q98CD7 rhizobium 1
45	90	7.0	429	17	Q9V0X8	Q9V0X8 pyrococcus

ALIGNMENTS

RESULT 1

Q8R7M2	ID	Q8R7M2	PRELIMINARY;	PRT;	255 AA.
AC	Q8R7M2;				
DT	01-JUN-2002 (TREMBLrel. 21, Created)				
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Putative transcriptional regulator, homologs of Bvg accessory factor.				
DE	factor.				
GN	TPE2381.				
OS	Thermoanaerobacter tengcongensis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;				
OC	Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.				
OX	NCBI_TaxID=119072;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MB4T / JCM11007;				
RX	MEDLINE=21992816; PubMed=11997336;				
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,				
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,				
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;				
RT	*A complete sequence of T. tengcongensis genome.*;				
RL	Genome Res. 12:689-700(2002).				
RL	EMBL; AE013180; AM25520.1;				
KW	Complete proteome; Hypothetical protein.				
SQ	SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;				

Query Match 64.5%; Score 825; DB 16; Length 255;
Best Local Similarity 62.0%; Pred. No. 2e-61;
Matches 158; Conservative 44; Mismatches 53; Indels 0; Gaps 0;

Qy	1	MLLVIDVGNNTNVLGIYDGERLVRDVRVSTDKARTDEYGLINELFRLAGLQDIQIRAV	60
Db	1	MLLAFDVGNTNVMGVFKGKLLHSFRISTDKNTYDEYGLVNLQIGYISLTIDDV	60
Qy	51	IISSVWPPLTGVLERLSLGVFGMPPLVVGPGIKTGMPTQYDNPREVGNADRVNAVGEK	120
Db	61	IISSVWPPLMTNLQVMSLKYFRTRKPIVVGPGIKTGINKIKYDNPKEVGNADRVNAVAYEL	120


```
Query Match          58.8%; Score 752; DB 16; Length 259;
Best Local Similarity 56.0%; Pred. No. 2.8e-55;
Matches 145; Conservative 47; Mismatches 63; Indels 4; Gaps 1;

QY 1 MLLVDVGNNTNVLGIDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 1 MLLVDVGNNTNVLGIDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 61 IISVVPLTGLVLERLSLGYFGMRPLVYVGPCKTGMPLOYDNPREVGNADRVNAVAGYEK 120
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 61 IISVVPLTGLVLERLSLGYFGMRPLVYVGPCKTGMPLOYDNPREVGNADRVNAVAGYEK 120
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 121 YRTSLIIVDFGTATTFDYNVRKGEYCGAIAPLGLVISTEALFORASKLPRVDIIRPSAII 180
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 121 HKKPMIIFDGTATTFCAITKEDYLGNGICPGIQISADALFERAALPRLELEKPKSVI 180
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 181 ARNTVNSMQAGIYGYVGLVDEIVTRMKAESK-----APRVATGGGLASLAPESKTEA 236
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 181 ARNTVNSMQAGIYGYVGLVDEIVTRMKAESK-----APRVATGGGLASLAPESKTEA 236
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 181 CKNTVTSMQAGIYGYVGLVDEIVTRMKAESK-----APRVATGGGLASLAPESKTEA 236
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 181 CKNTVTSMQAGIYGYVGLVDEIVTRMKAESK-----APRVATGGGLASLAPESKTEA 236
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 237 VEEYLTLEGLRIYERNR 255
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 237 VEEYLTLEGLRIYERNR 255
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 241 VDRKLTLEGLRIYERNR 259
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 241 VDRKLTLEGLRIYERNR 259
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

RESULT 5
Q97EB4 PRELIMINARY; PRT; 273 AA.
AC Q97EB4;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory
DE factor.
GN CAC3200.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=1466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007815; AAK81136.1; -
DR InterPro: IPR004619; Baf.
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;

Query Match          58.8%; Score 752; DB 16; Length 273;
Best Local Similarity 57.0%; Pred. No. 3e-55;
Matches 147; Conservative 48; Mismatches 59; Indels 4; Gaps 1;

QY 1 MLLVDVGNNTNVLGIDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 12 VILVDVGNNTNVLGIDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 61 IISVVPLTGLVLERLSLGYFGMRPLVYVGPCKTGMPLOYDNPREVGNADRVNAVAGYEK 120
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 72 IISVVPLTGLVLERLSLGYFGMRPLVYVGPCKTGMPLOYDNPREVGNADRVNAVAGYEK 120
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 121 YRTSLIIVDFGTATTFDYNVRKGEYCGAIAPLGLVISTEALFORASKLPRVDIIRPSAII 180
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
```

```
Db 132 YKRSLLIIFDGTATTFCAVRENGDYLGCAICPGIKVSEALFEKAALPRVELIKPAVAI 191
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
QY 181 ARNTVNSMQAGIYGYVGLVDEIVTRMKAESK-----APRVATGGGLASLAPESKTEA 236
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 192 CKNTISSQSGIVYGYIGQVRYIVRMKEELOEKEGEPLVYVATGGGLAKLISEEAKNDV 251
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
QY 237 VEEYLTLEGLRIYERNR 254
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 252 INPFLTLEGLRIYERNR 269
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

RESULT 6
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkut G.,
RA Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
DR EMBL: AL591974; CAD00748.1; -
DR MEROPS: M41.009; -
DR ListiList: LMO00221; -
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match          58.6%; Score 749; DB 16; Length 259;
Best Local Similarity 54.3%; Pred. No. 5e-55;
Matches 138; Conservative 56; Mismatches 60; Indels 0; Gaps 0;

QY 1 MLLVDVGNNTNVLGIDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 1 MLLVDVGNNTNVLGIDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 61 IISVVPLTGLVLERLSLGYFGMRPLVYVGPCKTGMPLOYDNPREVGNADRVNAVAGYEK 120
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 61 IISVVPLTGLVLERLSLGYFGMRPLVYVGPCKTGMPLOYDNPREVGNADRVNAVAGYEK 120
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 121 YRTSLIIVDFGTATTFDYNVRKGEYCGAIAPLGLVISTEALFORASKLPRVDIIRPSAII 180
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 121 YGTPVYVDFGTATTFYCIDESGYVQGGAIAPGIMISTEALYNRAALPRVDIAESSQII 180
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 181 ARNTVNSMQAGIYGYVGLVDEIVTRMKAESKAPRVATGGGLASLAPESKTEAVEEY 240
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
Db 181 GKSTVSSMQAGIFGVGCQCGEIIAEMKKQSNASPVVATGGGLARMIKTEKSAVDILDPF 240
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|

QY 241 LTLEGLRIYERNR 254
   :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:| :||:|||||:|||||:|
```

```

Db      241  LTLKGLLELYRRNK 254
RESULT 7
Q92F54  PRELIMINARY;      PRT;    259 AA.
AC      Q92F54;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Hypothetical protein lin0253.
GN      LIN0253.
OS      Listeria innocua.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC      Listeriaceae; Listeria.
OX      NCBI_TaxID=1642;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CLIP 11262 / SEROVAR 6A;
RX      PubMed=11679669;
RA      Glaser P., Franque L., Buchrieser C., Rusniok C., Amend A.,
RA      Raquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA      Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA      Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA      Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA      Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA      Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA      Madsen E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA      Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA      Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA      Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT      "Comparative genomics of Listeria species.";
RL      Science 294:849-852(2001).
DR      EMBL; AL596164; CAC95486.1; -
DR      ListList; LIN00253; -
DR      InterPro; IPR004619; Baf.
DR      Pfam; PF03309; Bvg_acc_factor; 1.
DR      TIGRFAMS; TIGR00671; baf; 1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match      58.2%; Score 745; DB 16; Length 259;
Best Local Similarity 54.3%; Pred. No. 1.le-54;
Matches 138; Conservative 54; Mismatches 62; Indels 0; Gaps 0;

QY      1  MLIVDVGNTNVLGIYDGERLVDRVSTDKARTTDEYGIINELFRAGLGLDQIRAV 60
      1  MLIVDVGNTNVTGVYKEQKLRHWRMTDRHRTSDELGMTVLNFFSTANLTPSDIQI 60
      61  IISVVPPLTGLTGLERLSLGYFGMRPLVVGPGIKTGMPIQYDNPREGADRIYNAVAYEK 120
      61  IISVVPPIHMAETMCVRYFNIRPLVPGIKTGLNKLVDNPREGSDRIYNAVAASEE 120
QY      121  YFTSLIIVDFGATTFDYNNRGEYCGGAIAPLVISTEALPQRASKLPVRDIIIPSII 180
      121  YGTPVIVDFGATTFYCYIDEAGYVGGGAIAPIGMISTEALYNRAAKLPVRDIAESSQII 180
QY      181  ARNTVNSMOAGIYGVVGLVDIVTRMKAESKDPRTATGGLASLAPESKTIEAVEEY 240
      181  GKSTVSMOAGIYGFYFGICEGIIAEMKKQSNTPVAVATGGLARMIYEKSSAVILDPF 240
QY      241  LTLGLRLIYERNR 254
      241  LTLKGLLELYRRNK 254
RESULT 8
Q9X8N6  PRELIMINARY;      PRT;    265 AA.
AC      Q9X8N6;
DT      01-NOV-1999 (TREMBLrel. 12, Created)
DT      01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE      Hypothetical protein SCO3380.
GN      SCO3380 OR SCE94.31C.
OS      Streptomyces coelicolor.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX      NCBI_TaxID=1902;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2);
RA      Oliver K., Harris D.;
RA      Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2);
RA      Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RA      Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RA      Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2);
RX      MEDLINE=97000351; PubMed=8843436;
RA      Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA      Kinashi H., Hopwood D.A.;
RT      "A set of ordered cosmids and a detailed genetic and physical map for
RT      the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL      Mol. Microbiol. 21:77-96(1996).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2) / M45;
RA      Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA      Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA      Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA      Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA      Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA      Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA      Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA      Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA      Hopwood D.A.;
RT      "Complete genome sequence of the model actinomycete Streptomyces
RT      coelicolor A3(2).";
RL      Nature 417:141-147(2002).
DR      EMBL; AL049628; CAB40880.1; -
DR      InterPro; IPR004619; Baf.
DR      Pfam; PF03309; Bvg_acc_factor; 1.
DR      TIGRFAMS; TIGR00671; baf; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match      49.2%; Score 629; DB 16; Length 265;
Best Local Similarity 51.1%; Pred. No. 6.3e-45;
Matches 134; Conservative 42; Mismatches 76; Indels 10; Gaps 5;

QY      1  MLIVDVGNTNVLGIYDGERLVDRVSTDKARTTDEYGIINELF---RLAG--LGLD 55
      1  MLTLDVGNTHFVLGLFDGEDIVHWRISTDSRRTADELAVLQGLMGHPLGLDELG-D 59
QY      56  QIRAVIISVVPPLTGLTGLERLSLGYFGMRPLY-VGPGIKTGMPIQYDNPREGADRIYNA 114
      60  GIDGIAICATVPESVLHELREVTTRYVDVPAVLVPEGVKTVPIITDHPKEVGADRIINA 119
QY      115  VAGYEKRTSLIIVDFGATTFDYNNRGEYCGGAIAPLVISTEALPQRASKLPVRDII 174
      120  VAAVELYGGPAIVVDFGATTFDAYSARGEYGGVIAPGIEISVEALYKGAOLRKIEVA 179
QY      175  RPSAIIARTVNSMOAGIYGVVGLVDIVTRMKAESKDP---RVATGGLASLAPES 231
      180  RPRSVIGKNTVEMOSGIVYGFAGVDGVNRMARELADDPDVTIVATGGLAPWVLGES 239
QY      232  RTIEAVEEYLTLEGLRLIYERN 253
      240  SVIDHEPWLTLMLGLRLIYERN 261
RESULT 9

```

```
Q9A6Z1          PRELIMINARY;          PRT;      261 AA.
ID Q9A6Z1;
AC Q9A6Z1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RC MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; Baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E6D07B0714EF5 CRC64;

Query Match          43.9%; Score 561; DB 16; Length 261;
Best Local Similarity 44.7%; Pred. No. 3.2e-39;
Matches 114; Conservative 50; Mismatches 89; Indels 2; Gaps 1;

QY 1 MLLVIDVGNNTNIVLGIYDGE---RLVDRVRSVTDKARTTDEYGLINELF-----RLAG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 MLLAEQNTNTFAIHGASVAQWRSATSTRTADEYVVLSQLSMQGLGFRAIDAV 61

QY 61 IISVVPPLTGVLERLSLGFGRPLVVGPIGTGMPIQVDNPREVGADRVNAVAGYEK 120
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 IISVVPQSTFNLRNLRSFNRYEPLVIGENAKLIGIDVRIEKPSEAGADRLNAIGAAMV 121

QY 121 YRSLIIVDFGTATTFDYVNRKGYCGGATAPGLVISTEALFORASKLPRVDIIRPSA-- 178
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 122 YGPLVVDSTGATTFDVAADGAFEGGIITAPGLNLSMQALHEAAKLPRIAIQRPAGNR 181

QY 179 IIAINTVNSMQAGIYGYVGLVDEIVTRMKAESKDAPRVITATGGLASLIAPESKTIKRAVE 238
   | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 182 IGVGDTVSAMQSGVFWGYISLIEGLVARIKAEKPEMTVIATGGVSLFEGATSDIDHFD 241

QY 239 EYTLLEGRLILYERN 253
   || : | : | : | : |
Db 242 SDLTIRGLLEIYRRN 256

RESULT 10
Q9CD56          PRELIMINARY;          PRT;      274 AA.
ID Q9CD56;
AC Q9CD56;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ML0232.
GN ML0232.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=TN;
RC MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL583917; CAC29740.1; -.
DR Leprona; ML0232; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; Baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match          40.9%; Score 522.5; DB 16; Length 274;
Best Local Similarity 40.2%; Pred. No. 5.9e-36;
Matches 109; Conservative 61; Mismatches 78; Indels 23; Gaps 4;

QY 1 MLLVIDVGNNTNIVLGIYDGE---RLVDRVRSVTDKARTTDEYGLINELF-----RLAG 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MLLAIDVRNTHVTVGLSSGSKHAKVQVQMRIRTESEVTADELAIIIDGLGDDSERLAG 60

QY 52 LGLDQIRAVIISVVPPLTGVLERLSLGFGRPLVVGPIGTGMPIQVDNPREVGADRI 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 -----AAALSTVPVSLHEVRIMLDQWPSVPHVLIETPGVTRTGPIPLVDNPKVEGADRI 113

QY 112 VNAVAGYEKYRTSLIIVDFGTATTFDYVNRKGYCGGAIAPGLVISTEALFORASKLPRV 171
   || : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 114 VNCLEAFHFGQAIVVDFGSSICVDVVSARKEFLGGAIPGVQVSSDAAARSALRRV 173

QY 172 DIIRPSAIIARNTVNSMQAGIYGYVGLVDEIVTRMKAESKD-----APRVITATGGLA 224
   : : || : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 174 ELAPRSVVGKNTVECHQAGVFGAGLVGLVGRMQDVEEFSGDLGNRVAVVATGHTA 233

QY 225 SLIAPESKTIKRAVEEYTLLEGRLILYERN 255
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 234 PLLLPETHVDVDRHLTLHGLRLVFERNRE 264

RESULT 11
O06282          PRELIMINARY;          PRT;      272 AA.
ID O06282;
AC O06282;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 29.3 kDa protein (transcriptional activator, putative,
DE Baf family).
GN RV3600C OR MTCY07H7B.22 OR MT3706.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton J., Squares R., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
```

complete genome sequence.":
 Nature 393:537-544(1998).

12] SEQUENCE FROM N.A.
 STRAIN-CDC 1551 / OSHKOSH;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z95557; CAB08944.1; -;
 DR EMBL: AE007170; AAK48063.1; -;
 DR TIGR: NT3706; -;
 DR TIGR: NT3706; -;
 DR TIGR: NT3706; -;
 DR InterPro: IPR004619; Baf.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGR: TIGR00671; baf; 1.
 DR TIGR: TIGR00671; baf; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 272 AA; 29304 MW; 5D70E6EF0F09AC8B CRC64;

Query Match 39.8%; Score 509.5; DB 16; Length 272;
 Best Local Similarity 39.6%; Pred. No. 7.2e-35;
 Matches 105; Conservative 65; Mismatches 82; Indels 13; Gaps 5;

QY 1 MLLVIDVGNNTVILVIGDYDGE-RLVDRWRVSTDKARTDEYGLINELFRLAGLGLDQ 56
 DB 1 MLLAIDVRNTHVTVGLSLGSKHAKVQOQRIKTESEVTALEUTDGLI---GEDSER 57
 QY 57 IRRAIVSSVVPPLTGTGLERLSLGYFGMRP-LVVGIGIKTGMPITQIDNPREVGGADRVNAV 115
 DB 58 LGCTAALSTVPSVLEHVRIMLQYWPSPVPHVLEPCVGTGIPLLVDNPKREVGGADRVNCL 117
 QY 116 AGYEKRTSLIIVDFGTATTFFVNRKGYCGGATAPGLVISTEALFORASKLPRVDIIR 175
 DB 118 AAYDRFKAIVVDFGSSICVDVWSAKGFEGLGATAPGVQVSSDAAARSALRRVELAR 177
 QY 176 PSAITARNVNSMQAGIYGYVGLVDEIVTRKAEKSKDAPRVATGTLASLAPESKTIE 230
 DB 178 PRSVGKNTVECMQAGVFGFAGLVGLGRIRNEDVSGFSDVDHVAIVATGHTAPLLDPE 237
 QY 231 SKTIEAVEYLTLEGLRILYERNR 255
 DB 238 LHTVDHYDQHLTQGLRLVFERNLE 262

RESULT 12
 QYRFEA

ID Q8RFE4 PRELIMINARY; PRT; 256 AA.
 AC Q8RFE4;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE Bvg accessory factor.
 GN FN0761.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25586;
 RX MEDLINE-21886394; PubMed-11889109;
 RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fontstein M., Kyriplides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586.";
 RT J. Bacteriol. 184:2005-2018(2002).
 DR EMBL: AE010586; AAL94957.1; -;

KW Complete proteome.
 SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 33.1%; Score 423; DB 16; Length 256;
 Best Local Similarity 35.5%; Pred. No. 1.2e-27;
 Matches 92; Conservative 60; Mismatches 99; Indels 8; Gaps 5;
 QY 1 MLLVIDVGNNTVILVIGDYDGE-RLVDRWRVSTDKARTDEYGLINELFRLAGLGLDQIRA 59
 DB 1 MIIGIDIGNHTVTVGLYDNGELISFTRIATNDKMTEDYFVFNITKYNISIKKVA 60
 QY 60 VISSVVPPLTGTGLERLSLGYFGMRPLVVGPGIKTGMPITQID---NPREVGGADRVNAV 116
 DB 61 ILISSVVPNIIITFFQFARKYKVEATIV--DLEKKLPFTFAKGINVTGFGADRIIDITE 118
 QY 117 GYEKY-RTSLIIVDFGTATTFFVNRKGYCGGATAPGLVISTEALFORASKLPRVDIIR 175
 DB 119 AMQKYPDKNLVIFDFGTATTYD-VLKKGVIYGGGILPGIDMSINALYGTAKLPRVKFTT 177
 QY 176 PSAITARNVNSMQAGIYGYVGLVDEIVTRKAEKSKDAPRVATGTLASLAPESKTIE 235
 DB 178 PSSVLGTDTKQIOQAIFFGYAGQIKHIKKINEELNEEIFVLATGGLGKILSAEIDEID 237
 QY 236 AVEEYLTLEGLRILYERNR 254
 DB 238 EYDANLSLKLTYLYKLNK 256

RESULT 13
 QYRFX54

ID Q9RX54 PRELIMINARY; PRT; 262 AA.
 AC Q9RX54;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE Hypothetical protein DR0461.
 GN DR0461.
 OS Deinococcus thermophilus.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE-20036896; PubMed-10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001905; AAF10040.1; -;
 DR TIGR: DR0461; -;
 DR InterPro: IPR004619; Baf.
 DR InterPro: IPR001220; Lectin_legB.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGR: TIGR00671; baf; 1.
 DR PROSITE: PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 KW SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 32.4%; Score 414.5; DB 16; Length 262;
 Best Local Similarity 36.8%; Pred. No. 6.6e-27;
 Matches 96; Conservative 50; Mismatches 102; Indels 13; Gaps 6;
 QY 2 LLLVIDVGNNTVILVIGDYDGE-RLVDRWRVSTDKARTDEYGLINELFRLAGLGLDQIRAV 60
 DB 6 LLAVIDGNTTVLGLADAGALTTHTWIRTNREMLPDDALQHLGLTLAGAPIP--RAA 63

```

QY 61 IISVVPPPLTGLVRLSLG---YFGMRPLVVGPGIKTGMPIQYDNPREVGDRIYNAVAG 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 VLSSVAPP---VGENYALKRHHMIDAFVAENLPDVTVELDPPGSGADRLCN-LFG 119
QY 118 YEKYTSL---IIVDFGTTATFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPVDII 174
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 AKYLGGLDYAVVDFGTTSTFDVVGRRFLGGLATGAQVSADALFAAAKLPRITIQ 179
QY 175 RPSAIARTNVSMQAGIYGYVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTI 234
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
180 APETAIGNVTWALQSLGVFGYAEVNDGILLRIRAEPLGCEAVAVATGFSRTVQICQEI 239
QY 235 EAVEEYLTLEGLRIYERNRE 255
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 DYYDETLLRLGLVELWASRSE 260

RESULT 14
083446 PRELIMINARY; PRT; 273 AA.
C O83446;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein TP0431.
GN TP0431.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC65417.1; -.
DR TIGR; TP0431; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 29.0%; Score 371.5; DB 16; Length 273;
Best Local Similarity 34.7%; Pred. No. 2.9e-23;
Matches 87; Conservative 56; Mismatches 101; Indels 7; Gaps 4;

QY 2 LLVIDVGNNTNVLGIY---DGERLVRD-WRVSTDKARTTDEYGIILNELFRLAGLQDI 57
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1. MLLIDVGNVSHVFGIQQNGRRVCVRELFRLAPDARKTQDEYSLIHALCERAGVGRASL 60
QY 58 RAVIISVVPPPLTGLVRLSLG---YFGMRPLVVGPGIKTGMPIQYDNP--REVGDRIYNAV 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 RDAFISVVPPVLTKTADAVAISGVQVFPVGPWAYEHLPLRIPPEVRAEIGTDLVANAV 120
QY 116 AGYERYRTSLIIVDFGTTATFDYVNRKGEYCGGATAPGLVISTEALFORASKLPVDIIR 175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 RAYVFRACVYVDCGTALTFTAVDGTGLIQGVAIAPGLTAVQSLHTGTGAQLPLVPLAL 180
QY 176 PSAIARTNVSMQAGIYGYVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTI 235
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
181 PDSVLGKDTHTAQQVGVVGRGLFVIRAMIAQQCKELGCRCAAVITGGLSRFSSE-VDFP 239
QY 236 AVEEYLTLEGL 246
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 240 PIDAQLTSLGL 250

RESULT 15
Q9WZY5 PRELIMINARY; PRT; 246 AA.
AC Q9WZY5;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
GN TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -.
DR TIGR; TM0883; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 28.3%; Score 361.5; DB 16; Length 246;
Best Local Similarity 36.8%; Pred. No. 1.7e-22;
Matches 91; Conservative 49; Mismatches 98; Indels 9; Gaps 6;

QY 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGIILNELFRLAGLQDI 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE---LFSLHPLLLGDAMREIKGI 57
QY 61 IISVVPPPLTGLVRLSLG---YFGMRPLVVGPGIKTG-MPIQYDNPREVGDRIYNAVAGYE 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 GVASVVPQNTVIERFSQKYFHISFIWV--KAKNGCVKNVKNPSEVGADRVANVAVFK 115
QY 120 KYRTSLIIVDFGTTATFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPYDIIRPSAI 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 EYKNGIILDMCTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLEVKKPADFV 174
QY 180 IARTNVSMQAGIYGYVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTI 239
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 VGGKTEENIRLGVVNGSVYALEGIIRIKVEYGDLP-VVLTGGQSKIVKDMIKH-EIFDE 232
QY 240 YLTLEGL 246
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
233 DLTIKGV 239

Search completed: June 24, 2003, 21:59:21
Job time : 22.7979 secs

```

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	88.5	6.9	463	4	US-09-134-001C-4160	Sequence 4160, Ap
2	86.5	6.8	719	4	US-08-765-907A-15	Sequence 15, Appl
3	86	6.7	330	4	US-09-134-001C-3582	Sequence 3582, Ap
4	83.5	6.5	388	4	US-09-222-938A-52	Sequence 52, Appl
5	83	6.5	4551	3	US-09-320-878-1	Sequence 1, Appl
6	77.5	6.1	1346	3	US-09-320-878-1	Sequence 4, Appl
7	77.5	6.1	1346	4	US-09-105-537-37	Sequence 37, Appl
8	77.5	6.1	11877	4	US-09-105-537-6	Sequence 6, Appl
9	77	6.0	514	4	US-09-134-001C-4347	Sequence 4347, Ap
10	77	6.0	553	2	US-08-663-566A-13	Sequence 13, Appl
11	77	6.0	553	2	US-08-484-575A-14	Sequence 14, Appl
12	77	6.0	553	2	US-08-023-610-13	Sequence 13, Appl
13	77	6.0	553	2	US-08-288-065A-13	Sequence 13, Appl
14	77	6.0	553	2	US-08-362-240A-13	Sequence 13, Appl
15	77	6.0	553	3	US-08-477-459-14	Sequence 14, Appl
16	77	6.0	553	3	US-08-479-869-14	Sequence 14, Appl
17	77	6.0	553	4	US-08-486-414-14	Sequence 14, Appl
18	77	6.0	553	4	US-08-804-372A-11	Sequence 11, Appl
19	77	6.0	553	5	PCT-US94-01826A-14	Sequence 14, Appl
20	77	6.0	553	5	PCT-US94-02252A-14	Sequence 14, Appl
21	77	6.0	553	5	PCT-US95-10245-13	Sequence 13, Appl
22	76.5	6.0	256	4	US-09-032-215-27	Sequence 27, Appl
23	76.5	6.0	683	4	US-09-620-412C-357	Sequence 357, App
24	76.5	6.0	948	4	US-09-556-877-194	Sequence 194, App
25	76.5	6.0	948	4	US-09-620-412C-194	Sequence 194, App
26	76.5	6.0	1776	4	US-09-556-877-179	Sequence 179, App
27	76.5	6.0	1776	4	US-09-620-412C-179	Sequence 179, App

APPLICANT: BMMAS-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
FILE OF INVENTION: Mutasynthesis
CURRENT APPLICATION NUMBER: US/08/765,907A
CURRENT FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 719
TYPE: PR1
ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-15

Query Match 6.8%; Score 86.5; DB 4; Length 719;
Best Local Similarity 25.0%; Pred. No. 0.34;
Matches 66; Conservative 35; Mismatches 108; Indels 55; Gaps 17;

QY 27 RVSTDKARTDEYGLINELFRLAGLGLDQIRAV-IISWVP-PLTGVLERLSLGYFGMR 84
DB 449 RLITD---GEIYEVCLTNMLRVP-RIDPLTAYRALRTVSPAPYAAVLOFPQATVLSSS 503

QY 85 P---LVVG-PGKTGMPIQYDNPREG--ADRVNA-VAGYKYRT-SLIIVDFGTATTF 136
DB 504 PERFLRIGADGWAESKPIKTRPGAGPAQDAAVKASLAAAEKDRSENLMIVDL----- 557

QY 137 DVVNRKGEYC--GGATAPGLV-ISTEALFORASKLPR-----VDIIRPSAIIA----- 181
DB 558 -VRNDLQVQCDIGSVHVPGLFEVEYATVHQLVSVVRGLAADVSRPRAVRAFPGGSM 616

QY 182 -RNTVNSM-----AGIYGYVGLVD-----EIVTRKAEKSDAPRVATGGIA 224
DB 617 GAPKVTMQLFIDRLEKPGRGVSGALYPALSGAADLSIVRTIVATEAATIGVGAVV 676

QY 225 SLIAPEKTIIE-AVEEYLTLEGLR 247
DB 677 ALSDDPDEVREMLLKAQTTLAALR 700

RESULT 3
US-09-134-001C-3582
Sequence 3582, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3582
LENGTH: 330
TYPE: PR1
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3582

Query Match 6.7%; Score 86; DB 4; Length 330;
Best Local Similarity 19.7%; Pred. No. 0.11;
Matches 63; Conservative 42; Mismatches 103; Indels 112; Gaps 15;

QY 1 MLLVIDVGNNTIVLGIYDGE-RLVRDNRVSTDKARTDEYGLINELFRLAGLGLDQIRA 59

DB 6 IILAADIGTCTCKLGTFDRDLQHLKWSIDTD---TSDHTGEL----- 45
QY 60 VIISVVVPLTGVLERLSLGYFGMRPLVVGPGIKTKMPLOYD-----NP 103
DB 46 -LLKNIYNSFTKIAEYKYDFNN---VVGVGIGVPGVDFDFTGVVYGVAVNLHWPDSVNV 100
QY 104 RE-----VGADRVNAVAGYKYR-----TSLIIVDFGTA----- 133
DB 101 REIFKQYVNCVYVVDNDANVAALGEKHKGAGBAGDADVAILTGLTGLGGGIIISNGEIVHGH 160
QY 134 -----TTFDYVNRKGEYCGG-----AIAPGLVISTEALFORAS-KLPRVDIIRP 176
DB 161 NGSAGAEIGHLRADFDQFQCGCKGSCIEIVASATGVNVLNVAIFYPKLTFKSSILQLIKD 220
QY 177 SAIIRNTVNSMQAGIYY-----GYVGLVDEIVTRKAEKSDAPRVATGG---LA 224
DB 221 NOVTAKEVDAKAGDQFCIFITEKRVANYIGYLCISIIS-----VTSNPKYIIVLGGGMSTA 275
QY 225 SLIAPEK-TIEAVEEYLT 243
DB 276 GLILIENT-----EYRNL 290

RESULT 4
US-09-222-938A-52
Sequence 52, Application US/09222938A
Patent No. 6437108
GENERAL INFORMATION:
APPLICANT: Youngman, Philip
APPLICANT: Fritz, Christian
APPLICANT: Murphy, Christopher
APPLICANT: Guzman, Luz-Maria
TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
FILE REFERENCE: 07334/060001
CURRENT APPLICATION NUMBER: US/09/222,938A
CURRENT FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 102
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 52
LENGTH: 388
TYPE: PR1
ORGANISM: Streptococcus pneumoniae
US-09-222-938A-52

Query Match 6.5%; Score 83.5; DB 4; Length 388;
Best Local Similarity 23.1%; Pred. No. 0.28;
Matches 51; Conservative 27; Mismatches 90; Indels 53; Gaps 9;

QY 28 VSTDKARTDEYGLINELFRLAGLGLDQIRAVIISVVVPLTGVLERL-SLGYFGMRPL 86
DB 147 IFTDTSALKEYPDLFKQYF-----AKLVPTDNKLAALNSAVWSGGTFI 191

QY 87 VVGPGIKTKMPLOYDNPREGADRVNAVAGYKYRTSLIIVDFGTATTFDYVNRKGEYC 146
DB 192 YVPGKGVKVDIPLQ-----TYFRINNENIG--QFERTLIIVDEGASV-----HYV 233

QY 147 GGAIAPGLVISTEALFORASKLPRVDIIRPSAIIARNTVNSMQAGIYYGVGLVDEIVTR 206
DB 234 ECTAP--TYSSNSLHAAIVEIFALD-----GAYMYTTIQNWSNDVY-----NLVTK 279

QY 207 MKAEKSDAPRVATGGLASLIAPEKTIIEAVEEYLTLEGLR 247
DB 280 RAKAKQDATVEWIDNGLGA-----KTMKYPSPVILDEGEAR 315

RESULT 5
US-09-320-878-1
Sequence 1, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.

APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4551
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-320-878-1

Query Match 6.5%; Score 83; DB 3; Length 4551;
Best Local Similarity 22.08; Pred. No. 17;
Matches 67; Conservative 38; Mismatches 108; Indels 92; Gaps 17;
QY 9 NTNVL-----GIDGERLVRD-----WVSTDKARTTDEYIGLINEFLRAG 51
DB 446 NAHVLEAPGVVEGASVESTVGSAGVGVVWVVSAAKSAALDA-----QIERLAA 499
QY 52 L-----GLD-----QTRAVIISVPLGVLRLSLGFTGMRPLVVGPIKGTGMPIQY 100
DB 500 FASDRDTGDGAVDAGAVDAGAVARVLAG-----GRAQFRAHVVGSG----- 545
QY 101 DNPVEGCA-----DRIVNAVAGYEKRTSLIIVDFG-----ATTED-----VYNRKGEY 145
DB 546 --PDLAALAAAEGLVGRVAGS-GVGRVAFVFGQGTQWAGMGAELLDSASVFAAAAE- 601
QY 146 CGGAAPGLVISTEALFORASKLP---RVDIIRP-----SALIARTVNS 187
DB 602 CEAALSPYDWSLEAVVRQAPGAPTLERVDVQVPTFAVMVSLARVQHHGVTQAVVGH 661
QY 188 MQAGIYYGYG---LVDEIVTRMKAESKD-APRVATGGLASLIAPESKTIIEAVEEYTL 243
DB 662 SOGEIAAAYVAGALSLDDAARVVTLRSKSAHAHLAGKGMISLALSEDAVLERLAGP--- 718
QY 244 EGLRI 248
DB 719 DGLSV 723

RESULT 6
US-09-320-878-4
Sequence 4, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538

EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1346
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-320-878-4

Query Match 6.1%; Score 77.5; DB 3; Length 1346;
Best Local Similarity 25.6%; Pred. No. 11;
Matches 32; Conservative 17; Mismatches 51; Indels 25; Gaps 5;
QY 146 CGGAAPGLVISTEALFORASKLP---RVDIIRP-----SALIARTVNS 187
DB 592 CETALSPYDWSLEAVVRQAPSAPTLDRVDVVQVPTFAVMVSLAKVQHHGVTPEAVIGH 651
QY 188 MQAGIYYGYG---LVDEIVTRMKAESKD-APRVATGGLASLIAPESKTIIEAVEEYTL 243
DB 652 SOGEIAAAYVAGALTLDAAARVVTLRSKSAHAHLAGKGMISLALSEATQRIE---NL 708
QY 244 EGLRI 248
DB 709 HGLSI 713

RESULT 7
US-09-105-537-37
Sequence 37, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 37
LENGTH: 1346
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-105-537-37

Query Match 6.1%; Score 77.5; DB 4; Length 1346;
Best Local Similarity 25.6%; Pred. No. 11;
Matches 32; Conservative 17; Mismatches 51; Indels 25; Gaps 5;
QY 146 CGGAAPGLVISTEALFORASKLP---RVDIIRP-----SALIARTVNS 187
DB 592 CETALSPYDWSLEAVVRQAPSAPTLDRVDVVQVPTFAVMVSLAKVQHHGVTPEAVIGH 651
QY 188 MQAGIYYGYG---LVDEIVTRMKAESKD-APRVATGGLASLIAPESKTIIEAVEEYTL 243
DB 652 SOGEIAAAYVAGALTLDAAARVVTLRSKSAHAHLAGKGMISLALSEATQRIE---NL 708
QY 244 EGLRI 248
DB 709 HGLSI 713

RESULT 8

US-09-105-537-6

; Sequence 6, Application US/09105537A

; Patent No. 6265202

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 6

; LENGTH: 11877

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-105-537-6

Query Match 6.1%; Score 77.5; DB 4; Length 11877;

Best Local Similarity 25.6%; Pred. No. 3.7e+02;

Matches 32; Conservative 17; Mismatches 51; Indels 25; Gaps 5;

QY 146 CGGAIAPGLVISTEALFORASKLP---RVDIIRP-----SALIARTVNS 187

DB 10842 CETALSPYVDWSLEAVVRQAPSAPTLDRVDVVPVTFAMVSLAKVQHHGITPEAVICH 10901

QY 188 MQAGIYGVG---LVDEIVTRKAEKSD-APRVATGGLASLIAPEKTIKTEAVEEYTL 243

DB 10902 SOGEIAAYVAGALTDDAARVVTLSRSIAHLAKGGMISLSEATQRIE---NL 10958

QY 244 EGLSI 248

DB 10959 HGLSI 10963

RESULT 9

US-09-134-001C-4347

; Sequence 4347, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4347

; LENGTH: 514

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4347

Query Match

Best Local Similarity 18.2%; Pred. No. 2.6;

Matches 63; Conservative 47; Mismatches 108; Indels 128; Gaps 13;

QY 10 TIVLIGYDGERLVRDVRVSTDKARTDEYGLINELFRLAGLDQIRAVIISVVPPL 69

DB 161 TVVLQYLY--ERFVIDYSLGSGTGMNLEQLDNDKALELGIERNQLPQLPHTVH--L 216

QY 70 TGVLERLSGYFGMRPLVVGPGIKTCMPFIQYDNPREGVAGDRIVNAVAGYKRTSLIIVD 129

DB 217 TGMKKRYA-----TLMGIDEQTPVIVGASDGVLSNLGVNSYQKEVAVT 260

QY 130 FGTA-----TTFDYVNRKGEY-CGGATAPGLVI-----STEALFORASKL 168

DB 261 IGTSGAIRTVINQPKTDEKGRIFCYILDKDQYVIGGPNNGGVWLRLRDEILASEVETA 320

QY 169 PRVDI-----IRPSA-----IIA 181

DB 321 KRLGVDPYDVLVTOIASRVKPGAEGLIFHPYLAGERAPLWADARGSFFGLTSLSHKHEMI 380

QY 182 RNTVNSMQAGIYGVVGLVDEIVTRKAEKSDAPRVI-ATGGLA-----NETPTIKATGGFASKSEIWRQMMADIFDFD 432

DB 381 RAALEGVLYNLYTVVLLALIEVM-----KTEAVEEYLTLE---GLRILYERNRE 255

QY 226 LIAPES-----KTEAVEEYLTLE---GLRILYERNRE 255

DB 433 LIVPESYESSCLGACVLGMKALGIDDFSVIKDMVGTTHAHPNQE 478

RESULT 10

US-08-663-566A-13

; Sequence 13, Application US/08663566A

; Patent No. 5853733

; GENERAL INFORMATION:

; APPLICANT: Cochran, Mark D

; APPLICANT: Macdonald, Richard D

; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys

; TITLE OF INVENTION: and Uses Thereof

; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: John P. White

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/663,566A

; FILING DATE: June 13, 1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P

; REGISTRATION NUMBER: 28,678

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)278-0400

; TELEFAX: (212)391-0526

; TELEX: 422523

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 553 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-663-566A-13

Query Match

Best Local Similarity 18.7%; Score 77; DB 2; Length 553;

Matches 49; Conservative 38; Mismatches 83; Indels 92; Gaps 11;

QY 41 ILINELFRLAGLDQIRAVIISVVPPLTGVLERLSGYFGMRPLVVGPGIKTCMPIOY 100

DB 233 LTIOALYNLAGNMDYL-----LTKLIGGNQLSSL-IGSGLITGNPILY 276

QY 101 DNPRE-VGADRIVNAVAGYKRTSLIIV-----DGTATTFYVNRKE----- 144

DB 277 DSOTQLLIGIQVTLPSVGNLNNMRATYLETLSVSTRGFASALVPKVVTRVGSVIELDTS 336

QY 145 -----YCGGAI-----APGL-----VISTEALFORAS-----KL 168

DB 337 YCIETDLDLYCTRIVTFPMSPGIYSLSGNTSACMYKTEGALTPPYMTIKGSVIANCEM 396

QY 169 PRVDIIRPSAIARTNVNSMQAGIYGVVGLVDEIVTRKAEKSDAPRVIATGGLASLIA 228

Db 397 TTCRCVNPPIISQN-----YGEAVSLID-----KQSCNVLSLGGITLRLS 437
Qy 229 PE-----SKTIEAVEEYLTLEG 245
Db 438 GEFDTYQKNISIQDSQVIITG 459

RESULT 11

US-08-484-575A-14
; Sequence 14, Application US/08484575A
; Patent No. 5925358
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran and David E. Junker
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,575A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELEPHONE: (212)278-0450
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-484-575A-14

Query Match 6.0%; Score 77; DB 2; Length 553;
Best Local Similarity 18.7%; Pred. No. 2.9;
Matches 49; Conservative 38; Mismatches 83; Indels 92; Gaps 11;
Qy 41 ILINELFRLAGLGDQIRAVIISVVPPLTGVLERLSLGYFGMRPLVVGPIKTMPIQY 100
Db 233 LTIOALYNLAGGNDYL-----LTKLGIGNQLSSL-IGSLITGNPILY 276
Qy 101 DNPRE-VGADRIVNAVAGYEKYRTSLIIV-----DFGTATTFDYNRKE----- 144
Db 277 DSQTQLLGIQVTLPSVGNLNNMRATYLETLSVSTRGFASALVPKVTVRVGSVIEELDTLS 336
Qy 145 -----YCGGAI-----APGL-----VISTEALFORAS-----KL 168
Db 337 YCIETDLDLCTRIVTFPMSPGIYSCLSGNTSACMSYKTEGALTTPYWTIKGSVIANCKM 396
Qy 169 PRVDIIRPSAIIARTVNSMQAGIYGYVGLVDEIVTRMKAESKDAPRVATGGLASLIA 228
Db 397 TTCRCVNPPIISQN-----YGEAVSLID-----KQSCNVLSLGGITLRLS 437
Qy 229 PE-----SKTIEAVEEYLTLEG 245
Db 438 GEFDTYQKNISIQDSQVIITG 459

RESULT 12

US-08-023-610-13
; Sequence 13, Application US/08023610

; Patent No. 5928648
; GENERAL INFORMATION:
; APPLICANT: Cochran Ph.D., Mark D
; APPLICANT: Macdonald Ph.D., Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/023,610
; FILING DATE: February 26, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-023-610-13
Query Match 6.0%; Score 77; DB 2; Length 553;
Best Local Similarity 18.7%; Pred. No. 2.9;
Matches 49; Conservative 38; Mismatches 83; Indels 92; Gaps 11;
Qy 41 ILINELFRLAGLGDQIRAVIISVVPPLTGVLERLSLGYFGMRPLVVGPIKTMPIQY 100
Db 233 LTIOALYNLAGGNDYL-----LTKLGIGNQLSSL-IGSLITGNPILY 276
Qy 101 DNPRE-VGADRIVNAVAGYEKYRTSLIIV-----DFGTATTFDYNRKE----- 144
Db 277 DSQTQLLGIQVTLPSVGNLNNMRATYLETLSVSTRGFASALVPKVTVRVGSVIEELDTLS 336
Qy 145 -----YCGGAI-----APGL-----VISTEALFORAS-----KL 168
Db 337 YCIETDLDLCTRIVTFPMSPGIYSCLSGNTSACMSYKTEGALTTPYWTIKGSVIANCKM 396
Qy 169 PRVDIIRPSAIIARTVNSMQAGIYGYVGLVDEIVTRMKAESKDAPRVATGGLASLIA 228
Db 397 TTCRCVNPPIISQN-----YGEAVSLID-----KQSCNVLSLGGITLRLS 437
Qy 229 PE-----SKTIEAVEEYLTLEG 245
Db 438 GEFDTYQKNISIQDSQVIITG 459
RESULT 13
US-08-288-065A-13
; Sequence 13, Application US/08288065A
; Patent No. 5961982
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Macdonald, Richard D
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys S-
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,240A
FILING DATE: Dec-22-94
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0400
TELEFAX: (212)391-0526
TELEX: 422523
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-240A-13

Query Match	6.0%;	Score 77;	DB 2;	Length 553;
Best Local Similarity	18.7%;	Pred. No. 2.9;		

QY		41	ILINELFRLAGLGDQIRAVIISVVPPITGVLERLSGYGMRPLVVGPIKTMPIQY	100
Db		233	LTIQALYNLAGNM DYL-----LTKLIGNQLSSL-IGSGLITGNPIL	276
QY		101	DNPRE-VGADRIVNAGYEKYTSLIV-----DFGTATTFDYYNRKE-----	144
Db		277	DSQTQLLGIOVLPSVGNLNMMRATYLETVLSVTRGFASALVPKVTVRVGSVIEELDT S	336
QY		145	-----YCGGAI-----APGL-----VISTEALFORAS-----KL	168
Db		337	YCETDLDIYCTRIVTFPMSPGIYSCLSGNTSACMTSKTEGALTTPYMTIKGSVIANCKM	396
QY		169	PWVDII RPSAIIARTNVSMQAQIYYGYVGLVDEIVTRMKAESKDAPRVITATGSLASLIA	228
Db		397	TTCRCVNPGLISQN-----YGEAVSLID-----KQSCNVLSLGGITURLS	437
		229	PE-----SKTIEAVEYLTL EG	245
Db		438	G EFDVTYQKNISIODSOVIITG	459

```

: GENERAL INFORMATION:
:
: APPLICANT: Cochran, Mark D
:
: APPLICANT: Junker, David
:
: APPLICANT: Wild, Martha A
:
: TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof
:
: NUMBER OF SEQUENCES: 60
:
: CORRESPONDENCE ADDRESS:
:

```

COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,459
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-459-14

Query Match 6.0%; Score 77; DB 3; Length 553;
Best Local Similarity 18.7%; Pred. No. 2.9;
Matches 49; Conservative 38; Mismatches 83; Indels 92; Gaps 11;
QY 41 ILINELFRAGLDQIRAVIISVVPLTGVLERLSLGYFCMRPLVVGPIKTCMPIQY 100
233 LTIQALYNLAGNMDYL-----LTKLGIGNNQLSSL-IGSLITGNPILY 276
101 DNPRE-VGADRIVANAGYKRYRTSLIIV-----DFGTATTFDYVNRKGE----- 144
Db 277 DSQTQLLGIVTLPSVGNLNNRATYLETLVSVSTRGFASALVPKVVTRVGSVIEELDT 336
QY 145 -----YCGGAI-----APGL-----VISTEALFORAS-----KL 168
Db 337 YCIETDLDLCTRIVTFPMSPGIYSCLSGNTSACMYSKTEGALTTPYMTIKGSVIANCKM 396
QY 169 PRVDIIRPSAIIARTVNSMQAGIYYGVYGLVDEIVTRMKAESKDAPRVATGGSLA 228
Db 397 TTCRCVNPPIGIIQN-----YGEAVSLID-----KQSCNVLSLGGITLRLS 437
QY 229 PE-----SKTIEAVEEYLTLEG 245
Db 438 GEFVTVQKNISIQDSQVIITG 459

Search completed: June 24, 2003, 21:49:19
Job time : 8.44934 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.6261 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-7
Perfect score: 1279
Sequence: 1 MLLVIDVGNNTNVLGIYDGE.....AVEEYLTLEGLRILYERNRE 255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

searched: 417779 seqs, 108206813 residues
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1279	100.0	255	9	US-09-813-453A-7
2	834	65.2	258	9	US-09-813-453A-49
3	785.5	61.4	256	9	US-09-813-453A-55
4	783	61.2	262	9	US-09-813-453A-45
5	771	60.3	254	9	US-09-813-453A-47
6	756	59.1	258	9	US-09-813-453A-2
7	650	50.8	233	9	US-09-813-453A-17
8	631.5	49.4	250	9	US-09-813-453A-3
9	629	49.2	265	9	US-09-813-453A-4
10	561	43.9	260	9	US-09-813-453A-51
11	558	43.6	219	9	US-09-813-453A-57
12	554	43.3	258	9	US-09-813-453A-6
13	509.5	39.8	272	9	US-09-813-453A-5
14	505.5	39.5	272	9	US-09-712-363-276
15	414.5	32.4	262	9	US-09-813-453A-8
16	371.5	29.0	273	9	US-09-813-453A-10
17	361.5	28.3	246	9	US-09-813-453A-9
18	338.5	26.5	212	9	US-09-813-453A-59
19	301	23.5	257	9	US-09-813-453A-53

20	266.5	20.8	262	9	US-09-813-453A-11	Sequence 11, Appl
21	208	16.3	244	9	US-09-813-453A-41	Sequence 41, Appl
22	172.5	13.5	241	9	US-09-813-453A-63	Sequence 63, Appl
23	163.5	12.8	249	9	US-09-813-453A-70	Sequence 70, Appl
24	159	12.4	249	9	US-09-813-453A-61	Sequence 61, Appl
25	148	11.6	267	9	US-09-813-453A-15	Sequence 15, Appl
26	143	11.2	460	9	US-09-813-453A-39	Sequence 39, Appl
27	141	11.0	248	9	US-09-813-453A-20	Sequence 20, Appl
28	141	11.0	592	9	US-09-813-453A-22	Sequence 22, Appl
29	139.5	10.9	229	9	US-09-813-453A-12	Sequence 12, Appl
30	139.5	10.9	592	9	US-09-813-453A-43	Sequence 43, Appl
31	128	10.0	242	9	US-09-813-453A-65	Sequence 65, Appl
32	126	9.9	257	9	US-09-813-453A-13	Sequence 13, Appl
33	100.5	7.9	223	9	US-09-895-913A-74	Sequence 74, Appl
34	100.5	7.9	223	9	US-09-813-453A-14	Sequence 14, Appl
35	100.5	7.9	223	9	US-09-813-453A-67	Sequence 67, Appl
36	94	7.3	610	10	US-09-815-242-11183	Sequence 11183, A
37	89	7.0	317	12	US-10-043-238-1	Sequence 1, Appl
38	89	7.0	317	12	US-10-043-238-3	Sequence 3, Appl
39	86	6.7	387	10	US-09-925-300-1477	Sequence 1477, Ap
40	83.5	6.5	388	9	US-10-154-251-52	Sequence 52, Appl
41	83	6.5	4551	9	US-09-793-708-1	Sequence 1, Appl
42	82	6.4	485	9	US-09-738-626-4551	Sequence 4551, Ap
43	81.5	6.4	714	10	US-09-815-242-10266	Sequence 10266, A
44	80.5	6.3	398	10	US-09-815-242-10390	Sequence 10390, A
45	80.5	6.3	472	10	US-09-822-863-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7
Query Match 100.0% Score 1279; DB 9: Length 255;
Best Local Similarity 100.0%; Pred. No. 7.3e-120;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYIGILINELFRLAGLGLDQIRAV 60
Db 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYIGILINELFRLAGLGLDQIRAV 60
Qy 61 IISVVVPLTGVLERLSLGYFGMRPLVVGPGIKTGMPYOYDNPREVGADRVNAVAGYEK 120
Db 61 IISVVVPLTGVLERLSLGYFGMRPLVVGPGIKTGMPYOYDNPREVGADRVNAVAGYEK 120
Qy 121 YRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIIRPSAI 180
Db 121 YRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIIRPSAI 180
Qy 181 ARNTVNSMOAGIYGYVGLVDEIVTRMKAESKADPRVIATGGLASLAPESKTI EAVEEY 240
Db 181 ARNTVNSMOAGIYGYVGLVDEIVTRMKAESKADPRVIATGGLASLAPESKTI EAVEEY 240

Db 181 ARNTVNSMOAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIIEAVEEY 240
QY 241 LTLEGLRILYERNRE 255
Db 241 LTLEGLRILYERNRE 255

RESULT 2
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

Query Match 65.2%; Score 834; DB 9; Length 258;
Best Local Similarity 62.7%; Pred. No. 2.2e-75;
Matches 160; Conservative 40; Mismatches 55; Indels 0; Gaps 0;
QY 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGIILNELFRLAGLGLDOIRAV 60
Db 1 MIFVLDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGIILNELFRLAGLGLDOIRAV 60
QY 61 IISVVVPLTGLVLERLSLGYFGMRPLVVGPGIKTGMPLOYDNPREVGADRIYNVAVGYEK 120
Db 61 IISVVVPMFALERMKCYFKIKPLVVGPGIKTGMDIKYDNPREVGADRIYNVAVAGIHL 120
QY 121 YRTSLIIVDFGATTDFYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAII 180
Db 121 YGSPLIIVDFGATTTCYINEKHQYMGGAIPGIMISTEALFARAAKLPRIEIRPDDII 180
QY 181 ARNTVNSMOAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIIEAVEEY 240
Db 181 GRNTVSAMOGIYGVGVGVEGIVSRMKAESKIPPKVIATGGLAPLASESDIIDVDPF 240
QY 241 LTLEGLRILYERNRE 255
Db 241 LTLEGLRILYERNRE 255

RESULT 3
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55
Query Match 61.4%; Score 785.5; DB 9; Length 256;
Best Local Similarity 60.9%; Pred. No. 1.5e-70;
Matches 156; Conservative 44; Mismatches 55; Indels 1; Gaps 1;
QY 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGIILNELFRLAGLGLDOIRAV 60
Db 1 MLLVFDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGIILNELFRLAGLGLDOIRAV 60
QY 61 IISVVVPLTGLVLERLSLGYFGMRPLVVGPGIKTGMPLOYDNPREVGADRIYNVAVGYEK 120
Db 61 IISVVVPMFALERMKCYFKIKPLVVGPGIKTGMDIKYDNPREVGADRIYNVAVAGIHL 120
QY 121 YRTSLIIVDFGATTDFYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAII 180
Db 121 YCAPSLIIVDFGATTTCYINEKHQYMGGAIPGIMISTEALFARAAKLPRIEIRPDDII 180
QY 181 ARNTVNSMOAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIIEAVEEY 239
Db 181 CKSTVSAMOGIYGVGVGVEGIVSRMKAESKIPPKVIATGGLAPLASESDIIDVDPF 240
QY 240 YLTLEGLRILYERNRE 255
Db 241 YLTLEGLRILYERNRE 256

RESULT 4
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match 61.2%; Score 783; DB 9; Length 262;
Best Local Similarity 57.3%; Pred. No. 2.8e-70;
Matches 145; Conservative 51; Mismatches 57; Indels 0; Gaps 0;
QY 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGIILNELFRLAGLGLDOIRAV 60
Db 1 MIFVLDVGNNTNVLGIYDGERLVRDWRVSTDKARTTDEYGIILNELFRLAGLGLDOIRAV 60
QY 61 IISVVVPLTGLVLERLSLGYFGMRPLVVGPGIKTGMPLOYDNPREVGADRIYNVAVGYEK 120
Db 61 IISVVVPMFALERMKCYFKIKPLVVGPGIKTGMDIKYDNPREVGADRIYNVAVAGIHL 120
QY 121 YRTSLIIVDFGATTDFYVNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAII 180
Db 121 YGSPLIIVDFGATTTCYINEKHQYMGGAIPGIMISTEALFARAAKLPRIEIRPDDII 180
QY 181 ARNTVNSMOAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIIEAVEEY 240

Db 181 GNTVSMQSGILYGVGVGIVKRMKEAKQPKVIATGGLAKLISESNVIDVDPF 240
QY 241 LTLEGLRIYERN 253
Db 241 LTKGLIYERN 253

RESULT 5

US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR FILING DATE: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 60.3%; Score 771; DB 9; Length 254;

Best Local Similarity 55.7%; Pred. No. 4.2e-69;

Matches 141; Conservative 52; Mismatches 60; Indels 0; Gaps 0;

QY 1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
Db 1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
QY 61 IISVVVPLTGVLERLSLGYFCGMRPLVVGPGIKTGMPTQYDNPKEVGDRIYNAVAGYEK 120
Db 61 IISVVVPMPSLEOMCKKXKPHVTPMIIIGPGIKTGLNKYDNPKEVGDRIYNAVAAIEL 120
QY 121 YRTSLIIVDFGTATTFDYNRKGECYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAI 180
Db 121 YCPAIVVDFTATTYCYLINEKKQYAGGVIAFGIMISTEALYHRAKLPRIEIAKPKQV 180
QY 181 ARNTVNSMOAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIBAVEEY 240
Db 181 GTNTIDSMQSGIFGVGVSDGVVVRMKAQAESEPKVIATGGLAKLIGTSETIDVIDSF 240
QY 241 LTLEGLRIYERN 253
Db 241 LTKGLIYERN 253

RESULT 6

US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR FILING DATE: 09/667,569
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 59.1%; Score 756; DB 9; Length 258;

Best Local Similarity 57.5%; Pred. No. 1.4e-67;

Matches 146; Conservative 46; Mismatches 62; Indels 0; Gaps 0;

QY 1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
Db 1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
QY 61 IISVVVPLTGVLERLSLGYFCGMRPLVVGPGIKTGMPTQYDNPKEVGDRIYNAVAGYEK 120
Db 61 IISVVVPMFALERMCTKYPHIEQIVGPGMKTGLNKYDNPKEVGDRIYNAVAAIHL 120
QY 121 YRTSLIIVDFGTATTFDYNRKGECYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAI 180
Db 121 YGNPLIVVDFTATTYCYIDENKQYMGGAIAPIIISTEALYHRAKLPRIEITRPNII 180
QY 181 ARNTVNSMOAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIBAVEEY 240
Db 181 GNTVSMQSGILFVGVEGIVKRMKWAQKDLKVIATGGLAPLIANESDCIDIVDPF 240
QY 241 LTLEGLRIYERN 254
Db 241 LTKGLIYERN 254

RESULT 7

US-09-813-453A-17

; Sequence 17, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE REFERENCE: OG2-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR FILING DATE: 2000-08-24

; PRIOR FILING DATE: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 17

; LENGTH: 233

; TYPE: PRT

; ORGANISM: Bacillus subtilis

US-09-813-453A-17

Query Match

Best Local Similarity 50.8%; Score 650; DB 9; Length 233;

Matches 124; Conservative 38; Mismatches 54; Indels 0; Gaps 0;

QY 1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
Db 1 MLLVIDGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
QY 61 IISVVVPLTGVLERLSLGYFCGMRPLVVGPGIKTGMPTQYDNPKEVGDRIYNAVAGYEK 120
Db 61 IISVVVPMFALERMCTKYPHIEQIVGPGMKTGLNKYDNPKEVGDRIYNAVAAIHL 120
QY 121 YRTSLIIVDFGTATTFDYNRKGECYCGGAIAPGLVISTEALFORASKLPRVDIIRPSAI 180
Db 121 YGNPLIVVDFTATTYCYIDENKQYMGGAIAPIIISTEALYHRAKLPRIEITRPNII 180

US-09-813-453A-4

```
RESULT 11
US-09-813-453A-57
; Sequence 57, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Dehalococcoides ethenogenes
US-09-813-453A-57

Query Match 43.6%; Score 558; DB 9; Length 219;
Best Local Similarity 49.8%; Pred. No. 6.6e-48;
Matches 104; Conservative 47; Mismatches 58; Indels 0; Gaps 0;

QY 2 LLVIDGNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAVI 61
D 1:|||||:|||||:|:|:| | ||| | :|:|:| |
Db 5 LVAVDIGNTSVNIGIFEKELLANWHLGSVAQRMADEVASLLGLLQHGHIPEELNVI 64
QY 62 ISSVVPPLTGVLERLSLGYFGMRPLVVG-PGKITGMPIDQYDNPREVGDRIVNAVAGYEY 121
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 65 MCSVVPPLTTTFEEFKSIFKAAPLVVGAGIKSGYKVRMDNPREVGADRIVNAARVLY 124
QY 122 RTSLLIVDFGTATTFDYVNRKGYCGGAIPGLVISTEALFORASKLPRVDIIRPSAIIA 181
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 125 PGACIIVDGNTATTFDISEGAYIGGAIAPGIATSAQIAEKTSKLPRIELIRPAKVI 184
QY 182 RNTVNSMOAGIYGVGLVDEIVTRMKAE 210
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 185 SNTVSAMQSGIYFGVIGLVEELVRRIQTE 213

RESULT 12
US-09-813-453A-6
; Sequence 6, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

Query Match 43.3%; Score 554; DB 9; Length 258;
Best Local Similarity 46.3%; Pred. No. 2.1e-47;

Matches 118; Conservative 41; Mismatches 94; Indels 2; Gaps 2;

QY 1 MLLVIDGNTNVLGIYDGERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQIRAV 60
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 1 MLLCIDCGNTNTVFSVWDGTFATWRIATDHRRTADEYFVWLNTLMQLKGL-QGRISEA 59
QY 61 IISVVVPLTGVLERLSLGYFGMRPLVVG-PGKITGMPIDQYDNPREVGDRIVNAVAGYE 119
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 60 IISSTAPRVVNLRLVLCNRYDCRPYVVGKPCGCELPVAPRVDPGTTVGPDLVNTVAGYD 119
QY 120 KYRSLIIVDFGTATTFDYVNRKGYCGGAIPGLVISTEALFORASKLPRVDIIRPSAI 179
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 120 RHGGDLIVDFGTATTFDVVAPDGCAYIGCVIAPGVNLSLEALHMAAALPHVDVTKPGCV 179
QY 180 IARTVNSMOAGIYGVGLVDEIVTRMKAEKSDAPRVATGGLASLIAPESKTEAVEE 239
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 180 IGTNTVACIQSGVYWGVIGLVEGIVQRIMERDRPMKVATGGLASLFDLGFDFDKVED 239
QY 240 YLTLEGLRILYERNR 254
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 240 DLTMHGLRLIFDYNK 254

RESULT 13
US-09-813-453A-5
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

Query Match 39.8%; Score 509.5; DB 9; Length 272;
Best Local Similarity 39.6%; Pred. No. 6.2e-43;
Matches 105; Conservative 65; Mismatches 82; Indels 13; Gaps 5;

QY 1 MLLVIDGNTNVLGIYDQ-----ERLVRDWRVSTDKARTTDEYGLINELFRLAGLGLDQ 56
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 1 MLLAIDVNTHTVVGLLSGMKERHAKVQVQWRITSEVTADELALTIDGLI---GEDSER 57
QY 57 IRAVIISVVVPLTGVLERLSLGYFGMRP-LVVGPGIKTGMPIQYDNPREVGDRIVNAV 115
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 58 LTGTAALSTVPSVLHEVRIMLDQYWPSPHVLIEPGRVIGPILLVDNPREVGADRIVNCL 117
QY 116 AGYEKYRTSLIIVDFGTATTFDYVNRKGYCGGAIPGLVISTEALFORASKLPRVDIIR 175
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 118 AAYDRFKAAIIVDFGSSICVDVVSAGKEFGLGAIAPGVQVSSDAAAARSALRRVELAR 177
QY 176 PSATARTVNSMOAGIYGVGLVDEIVTRMKAE-----SKDAP-RVATGGLASLIAP 230
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 178 PRSVGKNTVECMOAGAVFGVAGLVGDIREDVSGFSVDHDAIVATGHTAPLLE 237
QY 231 SKTTEAVEEYLTLEGLRILYERNR 255
D 1:|||||:|:| | ||| |||:|:| | ||| |||:|:| | ||| |||:|:| | |||
Db 238 LHTVDHYDQHLUTLOGLRLVFERNLE 262

RESULT 14
```


DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12292
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-276 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA75595.1; PID:g17133030; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3896

Query Match 10.0%; Score 127; DB 2; Length 276;
Best Local Similarity 32.1%; Pred. No. 0.0011;
Matches 53; Conservative 24; Mismatches 58; Indels 30; Gaps 9;

QY 96 YQRLGMDRWLAALGAFHLAKAGACLVLDLGTAAKADVFVSADG--EHLGGYICPGMPLMRSQ 153
DB 116 YPTLGLDRALALWAGMSGFPVLVIDAGTA--LTFTAADGKKNLVGGAILPGVGLQFAS 173
154 LRTHRRIRYDDASAEALSSLSPP--GRSTVEAVERGCVLMQGFAYTOLEQARVLMGEE 211
174 LGGQGTQLPQVEM---EAKSLPPRFALNTTAIQSGVI-----YTLIAGMRDFTTEW 223
QY 212 FTVF-----LTGGDAPL-----VRAALPQ--ARVV--PDLVFVGL 242
DB 224 LSLFPDGVKVAIRGGDRILLNLYLQALYPLDAAARLIVEPNLIEWGM 268

RESULT 15
RWVEV
genome polyprotein - equine arteritis virus
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: equine arteritis virus
A:Note: host Equus caballus (domestic horse)
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 29-May-1998
C:Accession: A39925; S10158; B39925
R:Den Boon, J.A.; Snijder, E.J.; Chirnside, E.D.; De Vries, A.A.F.; Horzinek, M.C.; Spaargaren, R.; Vries, A.A.F.; Chirnside, E.D.; Bredenoord, P.J.; Gravestien, L.A.; Horzinek, M.C.; Nucleic Acids Res. 18, 3241-3247, 1990
A:Title: All subgenomic mRNAs of equine arteritis virus contain a common leader sequence
A:Reference number: S10158; MUID:90287699; PMID:2162519
A:Accession: A39925
A:Molecule type: genomic RNA
A:Residues: 1-3175 <DEN>
A:Cross-references: EMBL:X53459
A:Note: a -1 ribosomal frameshift occurs between the codons AAC for 1727-Asn and CUG for 1728-Leu
R:De Vries, A.A.F.; Chirnside, E.D.; Bredenoord, P.J.; Gravestien, L.A.; Horzinek, M.C.; Nucleic Acids Res. 18, 3241-3247, 1990
A:Title: All subgenomic mRNAs of equine arteritis virus contain a common leader sequence
A:Reference number: S10158; MUID:90287699; PMID:2162519
A:Accession: A39925
A:Status: translation not shown
A:Molecule type: genomic RNA
A:Residues: 1-17 <VRI>
A:Cross-references: EMBL:X52277
C:Superfamily: equine arteritis virus RNA-directed RNA polymerase
C:Keywords: nucleotidyltransferase

Query Match 9.5%; Score 121; DB 1; Length 3175;
Best Local Similarity 21.9%; Pred. No. 0.056;
Matches 63; Conservative 37; Mismatches 94; Indels 94; Gaps 11;

QY 3 LEIDCGNSFKRWIRHVADAVIEGGGIVDSQQA-----LVAEVAALASVRLTGR 52
DB 91 LELLQHPAFAQLRVVDARLAEASVFISTDSAKRFPGARFALTPVYANAWVVSPPAN 150
QY 53 IVSVRSEETDLCALIAQAFVQAKVHPVREMAGVRNGYDDYQ-----RLGMDRW 104
DB 151 SLIVTTDQDQDGCFL-----KLLPPDRREAGRLRYNHYRQRTGWSLKTGLRW 201
QY 105 LAALG-----AFHLKAGC-----LVLDLGTAAKADFVSADGEH 138
DB 202 LGDLGLGINASSGGLKFHIMRSGPORAWHTTRSKLKYVYCDI---SEADWSCLPAGN 258

QY 139 LGGYICPGMPLMRSQLRTHRRIRYDDASAEALSSLSPPGRSTVEAVERGCVLMQGFAY 198
DB 259 YGGYNPPG-----DGAGGYECLAAMN-GATVVSA---GC----- 288
QY 199 TOLEQARVLMGEEFTVFLTGGDAPLVRAALPOARVVPDLVVFVGLAMAC 246
DB 289 -----SSDLWCDDDELAYRVFOLSPTFTTIPGRVCPNAY---AMIC 328

Search completed: June 24, 2003, 22:19:23
Job time : 11.2724 secs

A70955
 hypothetical protein RV3600c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70955
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; PMID:98295987; PMID:9634230
 A: Accession: A70955
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-272 <COL>
 A: Cross-references: GB:295557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113976
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: RV3600c
 C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 12.6%; Score 160.5; DB 2; Length 272;
 Best Local Similarity 24.4%; Pred. No. 1.3e-06;
 Matches 64; Conservative 43; Mismatches 106; Indels 49; Gaps 9;
 QY 1 MILEDCGNSFKRWIRHVADAVIEGGIVD-----SDQALVAEVAALASVRLTGCRIVS 55
 DB 1 MLLAIDVRNTH---TVVGLSGMKHAKVVQWQRIETSEVTADELATIDGLIG----- 52
 QY 56 VNSEETDLCALIAQAFVQAKVAVPREM-----AGVRNG-----YDDYQ 97
 DB 53 -----EDSERLTG--TAALSTVPSVLHEVRIMLDQYWPSPHYLIEPVRTGIPLLVDNPK 106
 QY 98 RLGMCDLAAALGAFHLAKGACLVIDLTAAKADFVSADGHEHGGVTCPCGMLRMSQLRTH 157
 DB 107 EVGADRVNCLAYDFRKAIVVDFGSSICVDVVSAGKEFLGALPAGVQVSSDAAAR 166
 QY 158 TRRIYDDASAEALSSLSGPRSTVEAVERGCVLMQLQGFAYTQLEQAR-----VLWGEF 212
 DB 167 SAALRRVELARPSV---VGKNTCEMQAGAVFGAGLVGLVGRIREDSVGSVDHV 222
 QY 213 TVFLTGDDAPLYRAALPQARVV 234
 DB 223 AIVATGHTAPLL---LPELHTV 241

RESULT 9
 E97293
 probable transcription regulator, homolog of Bvg accessory factor [Imported] - Clostridi
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C:Accession: E97293
 R: Noll, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A: Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
 A: Reference number: A96900; PMID:21359325; PMID:21359325
 A: Accession: E97293
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-273 <KUR>
 A: Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
 A: Experimental source: Clostridium acetobutylicum ATCC824
 C: Genetics:
 A: Gene: CAC3200
 C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 12.5%; Score 159; DB 2; Length 273;
 Best Local Similarity 23.6%; Pred. No. 1.7e-06;
 Matches 67; Conservative 49; Mismatches 92; Indels 76; Gaps 12;
 QY 1 MILEDCGNSFKRWIRHVADAVIEGGIVDSQALVAE-----VAALASVRLTGCRIVSV 56

DB 12 VILVLDVGN-----NIVLG---IYNDTKLTAEWRLSTOVLRSADVEYGIQVMNL 57
 QY 57 RSEETD-----ALCALIAQAFVQAKVAHPVREMAGVRNG-----GIKTGINKYD 94
 DB 58 FQDQKDDPTLVBEVVISSVVPNMVTSLEHMIRKRYKRNPLVVG-----GIRKTNIKYD 112
 QY 95 DYQRLGMDRWLAALGAFHLAKGACLVIDLTAAKADFVSADGHEHGGVTCPCGMLRMSOL 154
 DB 113 NPKEVGADRVNVAHAHEIKRSLIIDFGTATTCVRENGDYLGAICPGIKVSSEAL 172
 QY 155 RTHTRIRYDDASAEALSSLSPG-----RSTVEAVERGCVLMQLQGFAYTQLEQARVL--- 207
 DB 173 -----FEKAAKLPRVLIKPAYACKNTISSIQSGIV-----YGIQGVRIYVER 217
 QY 208 -----WGE-EFTVFLTGDDAPLYRAALPQARVV-PDLVFGVL 242
 DB 218 MKEELQEGEKEPLVVVATGGLAKLISEAKNVNDVINPFLTEGL 261

RESULT 10
 B87489
 transcription activator, probable Baf family [Imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: B87489
 R: Niemman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, N.J.; Emclaeava, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A: Title: Complete Genome Sequence of Caulobacter crescentus.
 A: Reference number: A87249; PMID:21173698; PMID:11259647
 A: Accession: B87489
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-261 <STO>
 A: Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
 C: Genetics:
 C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 12.1%; Score 153.5; DB 2; Length 261;
 Best Local Similarity 27.4%; Pred. No. 4.9e-06;
 Matches 66; Conservative 39; Mismatches 113; Indels 23; Gaps 8;
 QY 1 MILEDCGNSFKRWIRHVADAVIEGGIVDSQQA---LVAEVAALASVRLTGCR----- 52
 DB 2 MLLAIEQNTNTMFAIHGASVQWRSATSTRTADEYVWLSQLSQQGLGFRAIDAV 61
 QY 53 IVSVRSEETDLCALIAQAFVQAKVAHPVREMA---GVRNGYDDYQRLGMDRWLAALGA 110
 DB 62 ITSSVVPQSIFNRLNLSRRYFNVEPLV---IGENAKLGIDVRIEKPSEAGADRLVNAIGA 118
 QY 111 FHLAKGACLVIDLTAAKADFVSADGHEHGGVTCPCGMLRMSQLRTHTRIRYDDASAE 170
 DB 119 AMVYFGLVVDISGTTATFDIIVADGAFEGGIAPGINLSMQAL---HEAAAKLPRIATQ 176
 QY 171 ALLSLSPGSRSTVEAVERG-----VLMLQGFAYTQLEQARVLWGEETFTLTGCDAPLVRA 226
 DB 177 PAGNRIVGTDVTSAMQSGVFWYISLIEGLV-ARIKAER---GEPMTVIATGGVASLFG 232
 QY 227 A 227
 DB 233 A 233

RESULT 11
 F64627
 hypothetical protein HP0862 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000
 C:Accession: F64627
 R: Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzia, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V. A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: B81009
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-592 <NET>
A:Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733 A:Experimental source: serogroup B, strain MC58
A:Genetics:
C:Gene: NMB2075

Query Match 19.3%; Score 245; DB 2; Length 592;
Best Local Similarity 30.1%; Pred. No. 1.1e-13;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
3 LEIDCGNSFKRWIVHADVAVIEGGGIVDSQALVAEVAAL-----ASVRLTGCR 53
341 LLLDGGNSRLK-----AWVNGTFTATVGSAPYRDLSPGAEWAKEKADGNVIVGC-- 391
54 VSVRSEETDLCALTAQAFVQAKVAHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107
392 -----AVCGEFKKA-QVQEQARKIEWLPSQAQALGIRNHYRHPHEHGSORWFA 440
108 LGAFHLAKGACLVIDLTAAKADFVSADGEHLGGYICPGMPLMRSLRTHTRRI----- 161
441 LGSRRFSRNACVVVSCGTATVVDALDDGHYLGTTIMPGFHLMKESLAVRTANLNHACK 500
162 RYD-DASAEALSSLSPGSTVEAVERGCVLMQGFAYTQLEQARVLMGEFTVFLTGDD 220
501 RYFPFTTTGNVAVS-----GMMDAV-CGSVMVMHG-----RLKE-KTGACKPVDVITGGG 549
221 APLVRAALPQA-----RVVPLVFLVGL 242
550 AAKVAEALPPAFIAENTVRVADNLVIYGL 578

RESULT 3

H82031
Probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - *Neisseria meningitidis* N:Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C:Species: *Neisseria meningitidis*
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C:Accession: H82031
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel Holroyd, S.; Jegeris, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, ture 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: H82031
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-592 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737911 A:Experimental source: serogroup A, strain Z2491
A:Genetics:
C:Gene: birA; NMA0357
C:Keywords: ligase

Query Match 19.1%; Score 243; DB 2; Length 592;
Best Local Similarity 30.1%; Pred. No. 1.7e-13;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
3 LEIDCGNSFKRWIVHADVAVIEGGGIVDSQALVAEVAAL-----ASVRLTGCR 53
341 LLLDGGNSRLK-----AWVNGTFTATVGSAPYRDLSPGAEWAKEKADGNVIVGC-- 391
54 VSVRSEETDLCALTAQAFVQAKVAHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107

Db 392 -----AVCGEFKKA-QVQEQARKIEWLPSQAQALGIRNHYRHPHEHGSORWFA 440
QY 108 LGAFHLAKGACLVIDLTAAKADFVSADGEHLGGYICPGMPLMRSLRTHTRRI----- 161
Db 441 LGSRRFSRNACVVVSCGTATVVDALDDGHYLGTTIMPGFHLMKESLAVRTANLNHACK 500
QY 162 RYD-DASAEALSSLSPGSTVEAVERGCVLMQGFAYTQLEQARVLMGEFTVFLTGDD 220
Db 501 RYFPFTTTGNVAVS-----GMMDAV-CGSVMVMHG-----RLKE-KTGACKPVDVITGGG 549
QY 221 APLVRAALPQA-----RVVPLVFLVGL 242
Db 550 AAKVAEALPPAFIAENTVRVADNLVIHGL 578

RESULT 4

T36391
Hypothetical protein SCE94.31c - *Streptomyces coelicolor*
C:Species: *Streptomyces coelicolor*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36391
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T36391
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-265 <OLI>
A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCODEB:SCE94.31c A:Experimental source: strain A3(2)
C:Genetics:
C:Gene: SCODEB:SCE94.31c
C:Superfamily: *Streptomyces coelicolor* hypothetical protein SCE94.31c

Query Match 16.5%; Score 210; DB 2; Length 265;
Best Local Similarity 28.6%; Pred. No. 5.5e-11;
Matches 81; Conservative 46; Mismatches 90; Indels 66; Gaps 14;

QY 1 MILEDCGNSFKRWIVHADVAVIEGGGIVD-----SDQALVAEVAALASVRLTGCR 55
Db 1 MLITIDVGN-----HTVLGDFGCDIVEHWRISTDSRRTADELAVLGLMGH--P 51
QY 56 VSRSEETD-----ALCALTAQAFVQAKVAHPVREMA-----GVRNGY---- 93
Db 52 LLGDELGDGIDGATCATV-----PSVLHELREVTTRYGDPVAVLVEPGVKGTGPIL 104
QY 94 -DDYORLGRMDRWLAALGAFHLAKGACLVIDLTAAKADFVSADGEHLGGYICPGMPLMR 152
Db 105 TDHPKEVGADRIINAVALVELYGGPAIVVDFTATTFDVARSARGEYIGGVIAPIGIEISVE 164
QY 153 QLRTHTTRIRYDDASAEALSSLSPGSTVEAVERGCVLMQGFAYTQLE-----QARVL 207
Db 165 ALGVGAQLRKIEVAPRPSVI-----GKNTVEAMQSGIV--YGFA-GQVDGVVNRNAREL 216
QY 208 WG--BEFTVFLTGDDAPLVRALPQARV-----PDLVFLVGLAM 244
Db 217 ADDPDQDVTVIATGGLAPV---LGESSVIDEHPWLTLMGLRL 256

RESULT 5

H86937
conserved hypothetical protein ML0232 [imported] - *Mycobacterium leprae*
C:Species: *Mycobacterium leprae*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H86937
R:Colet, S.T.; Elgimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holroyd Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H86937

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.2724 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453A-61

Perfect score: 1269

Sequence: 1 MILELDCGNSFIKRWVHVA.....QARVVDLVFVGLAMACPLD 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	794.5	62.6	248	2 H83111	hypothetical prote
2	245	19.3	592	2 B81009	BirA protein/Bvg a
3	243	19.1	592	2 H82031	probable biotin-[a
4	210	16.5	265	2 T36391	hypothetical prote
5	176.5	13.9	274	2 H86937	conserved hypotet
6	168.5	13.3	242	2 A82637	conserved hypotet
7	163	12.8	267	2 I40327	baf protein - Bord
8	160.5	12.6	272	2 A70955	hypothetical prote
9	159	12.5	273	2 E97293	probable transcrip
10	153.5	12.1	261	2 B87489	transcription acti
11	144.5	11.4	223	2 F64627	hypothetical prote
12	137.5	10.8	223	2 G71807	hypothetical prote
13	139.5	10.2	233	2 S66100	conserved hypotet
14	127	10.0	276	2 A12292	hypothetical prote
15	121	9.5	3175	1 RRWVEV	genome polyprotein
16	118	9.3	257	2 S75559	hypothetical prote
17	118	9.3	273	2 D71326	conserved hypotet
18	117.5	9.3	246	2 D72320	conserved hypotet
19	114.5	9.0	209	2 H81382	hypothetical prote
20	114	9.0	254	2 F83660	hypothetical prote
21	109	8.6	259	2 AF1102	conserved hypotet
22	105	8.3	259	2 AF1464	conserved hypotet
23	98	7.7	638	2 A29440	signal recognition
24	97.5	7.7	517	2 G75593	uroporphyrin-III C
25	96.5	7.6	940	2 A87502	hypothetical prote
26	94.5	7.5	262	2 E75516	conserved hypotet
27	94.5	7.4	561	2 T36084	hypothetical prote
28	92.5	7.3	229	2 E70465	hypothetical prote
29	91	7.2	370	2 S27344	hupK protein - Rhi

30	91	7.2	629	2 E47096	membrane transloca
31	89	7.0	396	2 E84196	cell division prot
32	87.5	6.9	540	2 A26950	groEL2 protein - M
33	87.5	6.9	540	2 A43509	65K antigen mbaa -
34	87.5	6.9	2605	2 T18552	saframycin Mx1 syn
35	86	6.8	256	2 E83396	3-hydroxybutyrate
36	85	6.8	638	2 A24570	signal recognition
37	85	6.7	996	2 G87687	hypothetical prote
38	84.5	6.7	440	2 G87444	8-amino-7-oxononan
39	84	6.6	479	2 B87301	aldehyde dehydroge
40	84	6.6	562	2 E72608	probable hyuB APE1
41	84	6.6	767	2 E95924	probable mannose-1
42	82.5	6.5	305	2 AH3577	transcription regu
43	82.5	6.5	394	2 C87498	hypothetical prote
44	82.5	6.5	416	2 D75531	folyl-polyglutamat
45	82.5	6.5	428	2 D75582	hypothetical prote

ALIGNMENTS

RESULT 1

H83111

hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: H83111

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: H83111

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-248 <S>O>

A:Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4279

Query Match 62.6%; Score 794.5; DB 2; Length 248;

Best Local Similarity 63.5%; Pred. No. 2.4e-62;

Matches 158; Conservative 31; Mismatches 59; Indels 1; Gaps 1;

Qy 1 MILELDCGNSFIKRWVHVAADVIEGGIVDSQALVAEVAALASVRLTGRIYVSRSSE 60

Db 1 MILELDCGNSLIKRWVIEGAARSV-AGGLAESDDALVEQLTSQALPVRACRLVSRSEQ 59

Qy 61 ETDALCALIAQAFVQAKVAHPVREMAGVRNGYDDYQRLGMDRWLAALGAFHLAKGCLV 120

Db 60 ETSQLVARLEQLFPVSALVASSGGKLAGVRNGYLDYQRLGMDRWLAALGAFHLAKGCLV 119

Qy 121 IDLGTAAKADRVSGADGHELGICYCPGMPMSQLRTHRRIRYDDASAEALSSLSPPRS 180

Db 120 IDLGTAATSVLVAADGVHLGGYICPGMTLSQRTHRRIRYDDASAEALSSLSPPRS 179

Qy 181 TVEAVERGCVMILQGFAYTQLEQARVLWGEFTVLTGGDAPLVRAALPQARVVDLVFV 240

Db 180 TAEAVERGCVMILRGFVREQYAMACELLPDCEIFLTGGDAELVRDELAGARIMPDLFV 239

Qy 241 GLAMACPLD 249

Db 240 GLALACPTE 248

RESULT 2

B81009

BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: B81009

RA Bettencourt B.R.;
 RT "Rapid concerted evolution via gene conversion at the Drosophila heat
 RL shock protein Hsp70 genes";
 RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

[3]

RP SEQUENCE FROM N.A.

RX STRAIN=Berkley;

RC MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Branton R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R.,
 RA Gabor Miklos G.L., Abril J.F., Agbayani A., An H.J.,
 RA Andrews-Pfannkuch C., Baldwin D., Ballew R.M., Basu A., Baxendale J.,
 RA Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P.,
 RA Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck H.,
 RA Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H.,
 RA Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C.,
 RA Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z.,
 RA Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M.,
 RA Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C.,
 RA Ferraz C., Ferrier S., Fleischmann W., Foster C., Gabriellian A.E.,
 RA Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H.,
 RA Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J.,
 RA Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J.,
 RA Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z.,
 RA Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C.,
 RA Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J.,
 RA Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C.,
 RA McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C.,
 RA Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L.,
 RA Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K.,
 RA Nusskern D.R., Pacliet J.M., Palazzolo M., Pittman G.S., Pan S.,
 RA Pollard R.D., Puri V., Reese M.G., Reinert K., Remington K.,
 RA Saunders J.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I.,
 RA Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C.,
 RA Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R.,
 RA Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A.,
 RA Weinstein G.M., Weissbach J., Williams S.M., Woodage T.,
 RA Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S.,
 RA Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N.,
 RA Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W.,
 RA Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).

[4]

PRELIMINARY PARTIAL SEQUENCE FROM N.A.

RX MEDLINE=81076551; PubMed=6255408;

RA Toerok I., Karch F.

RT "Nucleotide sequences of heat shock activated genes in Drosophila
 melanogaster. I. Sequences in the regions of the 5' and 3' ends of
 the hsp 70 gene in the hybrid plasmid 56H8";
 RL Nucleic Acids Res. 8:3105-3123(1980).

[5]

SEQUENCE OF 1-100 FROM N.A.

RX MEDLINE=82197526; PubMed=6804941;

RA Ingolia T.D., Craig E.A.

RT "Drosophila gene related to the major heat shock-induced gene is
 transcribed at normal temperatures and not induced by heat shock";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:525-529(1982).

CC -1- INDUCTION: HEAT SHOCK INDUCES THE SYNTHESIS OF SEVEN PROTEINS AT
 FIVE OTHERWISE INACTIVE SITES IN THE POLYTENE CHROMOSOMES OF FRUIT
 FLY LARVAE. TWO SEPARATE SITES, PRODUCING TWO AND THREE COPIES,
 RESPECTIVELY, CODE FOR THE 70 KDA PROTEIN.

CC -1- MISCELLANEOUS: THERE ARE TWO COPIES OF THE GENE CODING FOR THIS
 PROTEIN AT CHROMOSOME LOCUS 87A7.

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: J01103; AAA28640.1; -
 DR EMBL: AF295939; AAG26893.1; -
 DR EMBL: AF295940; AAG26894.1; -
 DR EMBL: AF295941; AAG26895.1; -
 DR EMBL: AF295942; AAG26896.1; -
 DR EMBL: AF295943; AAG26897.1; -
 DR EMBL: AF295944; AAG26898.1; -
 DR EMBL: AF295945; AAG26899.1; -
 DR EMBL: AE003693; -; NOT_ANNOTATED_CDS.
 DR EMBL: V00213; CAA23495.1; ALT_SEQ.
 DR EMBL: V00214; CAA23496.1; ALT_SEQ.
 DR PIR: A03308; A03308.
 DR HSSP: P08107; LHJO.
 DR FlyBase: FBgn0013276; Hsp70Ab.
 DR InterPro: IPR001023; Hsp70.
 DR Pfam: PF00012; Hsp70.1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR PRODOM: PD000089; Hsp70.1.
 DR PROSITE: PS00297; Hsp70.1; 1.
 DR PROSITE: PS00329; Hsp70.2; 1.
 DR PROSITE: PS01036; Hsp70.3; 1.
 KW ATP-binding; Heat shock; Multigene family; Polymorphism.
 FT VARIANT 190 195
 FT VARIANT 195 195
 FT VARIANT 263 263
 FT VARIANT 490 490
 FT VARIANT 568 568
 FT VARIANT 574 574
 FT VARIANT 637 637
 FT CONFLICT 50 50
 FT CONFLICT 129 129
 FT CONFLICT 531 531
 FT CONFLICT 535 535
 FT CONFLICT 544 544
 SQ SEQUENCE 643 AA; 70265 MW; E9A2BF7F89905536 CRC64;
 Query Match 6.5%; Score 82; DB 1; Length 643;
 Best Local Similarity 22.7%; Pred. No. 10;
 Matches 58; Conservative 33; Mismatches 90; Indels 74; Gaps 12;
 Qy 19 VADAVIEGGIVDSQALVAEVAALASVRLTGCRIVSVRSEETDLCALIAQAFVQAK 78
 Db 137 ITDAVITVPAYFNDSDQATKDG-----HINAGLVRLINEPTAAAL-----AYGLD-- 184
 Qy 79 VAHPVRMAGVRNGYDDYQRLGMDRWLAALGAFHLAKG-----ACLVLDLGTAAKADFVSA 134
 Db 185 -----KNLKGERN-----VLFDLGGGTDFVSLITDEGSLFEVRSTAG 223
 Qy 135 DGEHLGG-----YICPGNPLRSOLRTHTRIRYDDASAEALSLSL 176
 Db 224 D-THLGGEDFNRLVTHLADEFKRY-----KKDLRSNPRALRLRLTAARAKRTLS 274
 Qy 177 PGRSTVBAVERGCVLMQLGGFAYTQEQARVLWGEEFTVFLTGDPALVRAALPOARV--- 233
 Db 275 --SSTEATIEDALFEGQDF-YTKVSRARF-----EELCADLFRNTLQPVKALNDKMDKG 328
 Qy 234 -VPDLVVFGLAMACP 247
 Db 329 QIHDIVLVGGSTRIP 343

Search completed: June 24, 2003, 22:11:57

Job time : 7.00886 secs

OX NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=122, AUS, B28, FRV3-1, QD18, and Z(H)1;
 RA Bettencourt B.R.;
 RT "rapid concerted evolution via gene conversion at the Drosophila heat
 RL shock protein hsp70 genes.";
 RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananthanades P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer J.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R.,
 RA Gabor Miklos G.L., Abril J.F., Agbayani A., An H.J.,
 RA Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J.,
 RA Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P.,
 RA Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J.,
 RA Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H.,
 RA Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C.,
 RA Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z.,
 RA Mays A.D., Dew I., Dietz S.M., Dodson K., Dunn P., Durbin K.J., Evans M.,
 RA Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C.,
 RA Ferraz C., Ferrier S., Fleischmann W., Foster C., Gabriellian A.E.,
 RA Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H.,
 RA Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J.,
 RA Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J.,
 RA Wei M.H., Iqbal M., Kalish J.M., Kalish F., Karpen G.H., Ke Z.,
 RA Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C.,
 RA Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J.,
 RA Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C.,
 RA McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C.,
 RA Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L.,
 RA Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K.,
 RA Nusskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S.,
 RA Pollard J., Puri V., Reese M.G., Reinert K., Remington K.,
 RA Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I.,
 RA Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C.,
 RA Stapleton M., Strong R., Sun X., Svirskaas R., Tector C., Turner R.,
 RA Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A.,
 RA Weinstock G.M., Weissbach J., Williams S.M., Woodage T.,
 RA Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S.,
 RA Zhan M., Zhang X., Zhao Q., Zheng L., Zheng X.H., Zhong F.N.,
 RA Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W.,
 RA Rubin G.M., Venter J.C.;
 RP "The genome sequence of Drosophila melanogaster.";
 RN Science 287:2185-2195(2000).
 CC -1- INDUCTION: HEAT SHOCK INDUCES THE SYNTHESIS OF SEVEN PROTEINS AT
 CC FIVE OTHERWISE INACTIVE SITES IN THE POLYTENE CHROMOSOMES OF FRUIT
 CC FLY LARVAE. TWO SEPARATE SITES, PRODUCING TWO AND THREE COPIES,
 CC RESPECTIVELY, CODE FOR THE 70 KDA PROTEIN.
 CC -1- MISCELLANEOUS: THERE ARE TWO COPIES OF THE GENE CODING FOR THIS
 CC PROTEIN AT CHROMOSOME LOCUS 87A7.
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF295933; AAG26887.1; -;
 DR EMBL; AF295934; AAG26888.1; -;
 DR EMBL; AF295935; AAG26889.1; -;
 DR EMBL; AF295936; AAG26890.1; -;

DR EMBL; AF295937; AAG26891.1; -;
 DR EMBL; AF295938; AAG26892.1; -;
 DR EMBL; AE003693; AAG22148.1; ALT_SEQ.
 DR HSP; P08107; 1HJO.
 DR Flybase; Fgn0013275; Hsp70Aa.
 DR InterPro; IPR001023; Hsp70; 1.
 DR Pfam; PF00012; Hsp70; 1.
 DR PRINTS; PR00301; HEATSHOCK70.
 DR PRODOM; PD000089; Hsp70; 1.
 DR PROSITE; PS00297; Hsp70.1; 1.
 DR PROSITE; PS00329; Hsp70.2; 1.
 DR PROSITE; PS01036; Hsp70.3; 1.
 KW ATP-binding; Heat shock; Multigene family;
 KW Polymorphism.
 FT VARIANT 441 441 Y -> S (IN STRAIN FRV3-1).
 FT VARIANT 495 495 K -> T (IN STRAIN QD18).
 FT VARIANT 593 593 L -> M (IN STRAIN 122).
 FT VARIANT 615 615 MISSING (IN STRAIN 122).
 FT VARIANT 636 636 P -> R (IN STRAIN QD18).
 FT CONFLICT 241 241 E -> D (IN REF. 1).
 FT CONFLICT 350 350 E -> D (IN REF. 1).
 SQ SEQUENCE 642 AA; 70188 MW; A2COE3030CECE32 CRC64;
 Query Match 6.5%; Score 82; DB 1; Length 642;
 Best Local Similarity 22.7%; Pred. No. 9.9;
 Matches 58; Conservative 33; Mismatches 90; Indels 74; Gaps 12;
 QY 19 VADAVTEGGIVDSQALVAEVAALASVRLTGCRIVSVRSEETDCAICAAFAVQAK 78
 DB 136 ITDAVTVPAYFNDQSQRATKDG-----HAGLVLRIRINEPTAAL-----AYGLD-- 183
 QY 79 VAHPVREMGVRNGYDDYQRLGMDRWLAALGAFHLAGK-----ACLVIDLTGAADVFSA 134
 DB 184 -----KNLKGKRN-----VLIFDGGTDFVDSILTIDGSLFEVRSSTAG 222
 QY 135 DSEHLGG-----YICGMPLMRSLQTHPRIRYDDASERALSLS 176
 DB 223 D--THLGGEDFDNRLVTHLAEFKRY-----KKDLRSPRALRLRTAAERAKRTLS 273
 QY 177 PGRSTVEAVERGCVLMQGFAYTQLPQARVLMGEFTVLTGGDAPLVRAALPQARV--- 233
 DB 274 --SSTEATIEIDALFEGQDF-YTKVSRARF---EELCALDFRLTLPQVEKALNDKMDKG 327
 QY 234 -VPDLVFGVGLAMACP 247
 DB 328 QIHDIVLVGGSTRIP 342
 RESULT 15
 HS71_DROME STANDARD; PRT; 643 AA.
 AC P02825;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Major heat shock 70 kDa protein Ab (Heat shock protein 70Ab) (HSP70-87A7).
 GN HSP70AE OR HSP70A.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82055575; PubMed=6795353;
 RA Karch F., Toeroker I., Tissieres A.;
 RT "Extensive regions of homology in front of the two hsp70 heat shock
 RL variant genes in Drosophila melanogaster.";
 RN J. Mol. Biol. 148:219-230(1981).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=122, 58H8, AUS, B28, FRV3-1, QD18, and Z(H)1;


```
DR TIGR: CC2008;
DR InterPro: IPR002819; HD
DR InterPro: IPR003507; ME_pplase_HDC.
DR Pfam: PF01966; HD; 1.
DR SMART: SM00471; HDC; 1.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 394 AA; 44624 MW; A1010E7C2DD2955 CRC64;

Query Match
Best Local Similarity 6.5%; Score 82.5; DB 1; Length 394;
Matches 54; Conservative 26; Mismatches 74; Indels 101; Gaps 11;

QY 71 QAEAVQAKVHPVREMAGV-----RNG-----YDDVQRLGMD 102
DB 132 QTFRVVTELEHRYDPDFIGLNTWTETLEGVIKHNGPVTNKLKPKSWKASKYDNEYELGLG 191
QY 103 RWLAALGAFHLAKGACLVLDIGTAADKADFSADGHEHLGGYICPGM-----PLMRSQ 153
DB 192 TWASA-----EAQVAALADDIAYNNHVDVDDGVYTAGLFTLDDLMDVPLIGPI 237
QY 154 LR-----THPRRIR-----YDASAE---RALSSLSPGRSTVEAVERGCVL 191
DB 238 LAAVKSRPDLDAHLTHLEAVRRMIGAMVDDVMGETLHRAAASGVQSADDDVRLDHALVA 297
QY 192 M-----LOGFAYTQL-----EQARVLNGEFTVLTGGDA-PLYRAALPQ 230
DB 298 FSSDMAEDLAKRGLFELRYHRHVRNTRNSQARKILGEMFALFLREPEVLPTVWFASQ 357
QY 231 -----ARVVPDLV 238
DB 358 NRDEAGRARVVCVDI 372

RESULT 10
IMDH_TRYBB
ID IMDH_TRYBB STANDARD; PRT; 512 AA.
AC P50098;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205) (IMP
DE dehydrogenase) (IMPDPH) (IMPD).
OS Trypanosoma brucei brucei.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EA164 / ISTAT 1.7;
RC MEDLINE=95050714; PubMed=7961861;
RA Wilson K., Berens R.L., Sifri C.D., Ullman B.;
RT "Amplification of the inosinate dehydrogenase gene in Trypanosoma
RT brucei gambiense due to an increase in chromosome copy number.";
RL J. Biol. Chem. 269:28979-28987(1994).
CC -!- FUNCTION: IMP IS THE RATE LIMITING ENZYME IN THE DE NOVO SYNTHESIS
CC OF GUANINE NUCLEOTIDES AND THEREFORE IS INVOLVED IN THE REGULATION
CC OF CELL GROWTH.
CC -!- CATALYTIC ACTIVITY: Inosine 5'-phosphate + NAD(+) + H(2)O -
CC xanthosine 5'-phosphate + NADH.
CC -!- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -!- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDPH AND TO
CC GMP REDUCTASE.
CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR HSSP: P12268; 1B30.

DR InterPro: IPR000644; CBS domain.
DR InterPro: IPR003009; FMN_enzyme.
DR InterPro: IPR001093; IMPDH/GMPRTase.
DR Pfam: PF00478; IMPDH_C; 1.
DR Pfam: PF00571; CBS; 2.
DR Pfam: PF01574; IMPDH_N; 1.
DR SMART: SM00116; CBS; 2.
DR TIGRPFAMS: TIGR01302; IMP_dehydrog; 1.
DR PROSITE: PS00487; IMP_DH_GMP_RED; 1.
KW Oxidoreductase; NAD; GMP biosynthesis; Purine biosynthesis; Repeat;
KW CBS domain.
FT DOMAIN 108 163 CBS 1.
FT DOMAIN 170 226 CBS 2.
FT BINDING 325 325 IMP (POTENTIAL).
SQ SEQUENCE 512 AA; 55708 MW; 1A86C46AE6445045 CRC64;

Query Match
Best Local Similarity 6.5%; Score 82.5; DB 1; Length 512;
Matches 58; Conservative 32; Mismatches 95; Indels 43; Gaps 11;

QY 1 MILEDCGN-----SFIKRVIHVADAVIEGGIVDSQALVAEVAALASVRL-TGCRIV 54
DB 265 LVLDSSQGTIIYQVSVFIRWVKTYPHLEVVAGNVVTQDAKNLIDAGADSLRIGMSGSI 324
QY 55 SVRSEETDLCALIAQAFVAQVAH-----PVREMGVNRNGYDDYQRLGMDRWLAA 107
DB 325 CITQE-----VLACGRPQATAY-KVARYAASRGVPCVADGGLRNVGDVCKALAVGANVAM 379
QY 108 LGAFHLAKGACLVLDLGTGTA-AKADFSVADGHEHLGGYICPGMPLMRSLRTHTRIRYDDA 166
DB 380 LGS-----MIAGTSETPGYFFKDGMRKGY--RGMGSIDAMLQGR-----E 419
QY 167 SAERALS---SLSPGRSTVEAV-ERGCVMILQGFAYTOLEQARVLWGE 210
DB 420 SGKRYLSENETLQVAGVAGVLDKGSVLKLLAYTHKGLQGSADIDGE 467

RESULT 11
ILVC_MYCLE
ID ILVC_MYCLE STANDARD; PRT; 333 AA.
AC O33114;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
DE isomerase) (ALPHA-keto-beta-hydroxyacil reductoisomerase).
GN ILVC OR ML1694 OR MLCB637.22.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RC MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Elgmeier K., Farkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
CC - (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -!- PATHWAY: valine and isoleucine biosynthesis; second step.
CC -!- SIMILARITY: BELONGS TO THE KETOL-ACID REDUCTOISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

DR TIGRFAMS: TIGR00475; seib; 1.
 DR PROSITE; PS00301; EFATOR-GTP; 1.
 KW Protein biosynthesis; GTP-binding.
 FT NP_BIND 10 17 GTP (BY SIMILARITY).
 FT NP_BIND 60 64 GTP (BY SIMILARITY).
 FT NP_BIND 115 118 GTP (BY SIMILARITY).
 SQ SEQUENCE 634 AA; 70666 MW; 626986A693A8296 CRC64;
 Query Match
 Best Local Similarity 6.8%; Score 86; DB 1; Length 634;
 Matches 57; Conservative 21; Mismatches 82; Indels 50; Gaps 10;
 QY 28 GIVDSQALVAEVAALASVRLTGCRIVSRSEETDLCALTAQAFAVQAKVAHPVREMA 87
 DB 138 GTVLEDAPIV-EVSA-----LTGEGIAELR--EQLDALAAVT-----PPRFAA 177
 QY 88 GVRNGYDDYQRLGMDRWLAALGAFHLAKGACLVLDLGTAAKADFYSDGHEGLGGYICPGM 147
 DB 178 G-----RVRLPIDRVFSTGTVTG---TLMSGTIKVGDELEVQPE-----GL 219
 QY 148 PLMRSQLRTHRRIRYDDASAEALSSLSPGSTVEAVERGCVLMQGF-----AY 198
 DB 220 KTRARNLVGHGRTVK--EARQORVAVNLGIET-EAVHRGSSLLTPGFLTPTYRLDASF 276
 QY 199 TQLEQARVLWGEFTVLTGDPAPLVRAAL 228
 DB 277 KLLNGARPLANDRVHFLGTSEALGRVVL 306
 RESULT 8
 SRPR_CANFA STANDARD; PRT; 638 AA.
 AC F06625;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Signal recognition particle receptor alpha subunit (SR-alpha)
 DE (Docking protein alpha) (DP-alpha).
 GN SRPR.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86065454; PubMed=2999608;
 RA Lauffer L., Garcia P.D., Harkins R.N., Coussens L., Ullrich A.,
 RA Walter P.;
 RT "Topology of signal recognition particle receptor in endoplasmic
 reticulum membrane.";
 Nature 318:334-338(1985).
 CC -1- FUNCTION: THIS INTEGRAL MEMBRANE PROTEIN ENSURES, IN CONJUNCTION
 WITH SRP, THE CORRECT TARGETING OF THE NASCENT SECRETORY PROTEINS
 TO THE ENDOPLASMIC RETICULUM MEMBRANE SYSTEM.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM MEMBRANE. THOUGHT TO
 BE ANCHORED IN THE MEMBRANE THROUGH AN INTERACTION WITH SR-BETA,
 WHICH CONTAINS A BONA FIDE TRANSMEMBRANE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; X03184; CAA36945.1;
 DR PIR; A24570; A24570.
 DR HSSP; O07347; 1FFH.
 DR InterPro: IPR000897; SRP54.
 DR Pfam; PF00448; SRP54; 1.

DR Pfam; PF02881; SRP54_N; 1.
 DR PROSITE; PS00819; SRP54; 1.
 DR PROSITE; PS00300; SRP54; 1.
 KW Signal recognition particle; Transmembrane; Receptor;
 KW Endoplasmic reticulum; GTP-binding;
 FT TRANSMEM 11 33 POTENTIAL.
 FT NP_BIND 425 432 GTP (BY SIMILARITY).
 FT NP_BIND 520 524 GTP (BY SIMILARITY).
 FT NP_BIND 588 591 GTP (BY SIMILARITY).
 SQ SEQUENCE 638 AA; 69672 MW; AD88CE554ACC2BC6 CRC64;
 Query Match
 Best Local Similarity 6.8%; Score 86; DB 1; Length 638;
 Matches 64; Conservative 31; Mismatches 82; Indels 66; Gaps 14;
 QY 33 QDALVAEVAALASVRLTGCRIVSRSEET-----DALCALTAQAFAVQAKVAHPVREMA 87
 DB 347 DHLIAKNVAADIAVLQ--CESVANKLEGKVMGTFTSTVTKQALQESILVQLQQRVD 404
 QY 88 GVRNGYD-DYOR-----LGMDRWLAALGAFHLAKGACLVLDLGTAAKADFYSA 134
 DB 405 MLRDINDAQRHQRPYVTVTCGVNGYGVKSTNLAKISFWLLENGFSVLI-----AACDTFRAG 460
 QY 135 DGEHLGGYICPGMPLMRSQLRTHRRIRYDDASAEALSSLSLSP-----GRSTVEAVERGC 189
 DB 461 AVEH-----VRTHRR-----LSALHPKHKAGPTWVQFEKGY 494
 QY 190 VLMLQGFAYTQLEQARVLWGEFTVFL--TGG-----DAPLVRAALPQARV-VPDLV-FVG 241
 DB 495 GKDAAGIAEATAFAR---NQGFVVLVDTAGRMQDNAPLMTALAKLITVNTPTDLVLFVG 551
 QY 242 LAM 244
 DB 552 EAL 554
 RESULT 9
 DGTL_CAUCR STANDARD; PRT; 394 AA.
 ID DGTL_CAUCR
 AC Q9A6S5;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Deoxyguanosinetriphosphate triphosphohydrolase-like protein.
 GN CC2008.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy K.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus.";
 Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -1- SIMILARITY: BELONGS TO THE DGTPASE FAMILY. SUBFAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; AF005874; AAK23983.1;

"All subgenomic mRNAs of equine arteritis virus contain a common leader sequence.";
 RT Nucleic Acids Res. 18:3241-3247(1990).
 CC -1- FUNCTION: RNA-DIRECTED RNA POLYMERASE & POSSIBLE HELICASE. A
 CC ROLE FOR NTP-BINDING PROTEINS IN RNA DUPLEX UNWINDING HAS BEEN
 CC SUGGESTED. ALSO CONTAINS A PROTEASE DOMAIN.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -1- MISCELLANEOUS: A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS
 CC FOR 1727-ASN AND 1728-LEU.
 CC -1- SIMILARITY: WITH THE POLYMERASE PROTEIN OF OTHER CORONAVIRUSES AND
 CC OF TOROVIRUSES.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S32.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X53459; CAA37539.1; ALT_SEQ.
 CC EMBL; X53459; CAA37540.1; ALT_SEQ.
 CC EMBL; X52277; CAA36520.1; --
 CC PIR; S10158; S10158.
 CC PIR; A39925; RRVVEV.
 CC MEROPS; C31.UNW; --
 CC MEROPS; C32.001; --
 CC MEROPS; C33.001; --
 CC MEROPS; S32.001; --
 CC InterPro: IPR000606; Viral_helicase1.
 CC Pfam: PF01443; Viral_helicase1.
 CC RNA-directed RNA polymerase; Transferase; Helicase; ATP-binding;
 CC Hydrolase; Serine protease; Zinc-finger.
 CC CHAIN 1 1727 ORFLB.
 CC FT CHAIN 1728 3175 TRYPsin-LIKE SERINE PROTEASE.
 CC FT DOMAIN 1080 1220 HELICASE.
 CC FT DOMAIN 1218 1506 HELICASE.
 CC FT DOMAIN 2098 2306 POLYMERASE.
 CC FT ACT_SITE 1103 1103 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 1129 1129 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 1184 1184 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ZN_FING 2368 2414 BY SIMILARITY.
 CC FT NP_BIND 2528 2535 ATP (BY SIMILARITY).
 CC FT SEQUENCE 3175 AA; 345275 MW; EDC8BA0E31DE695 CRC64;

Query Match 9.5%; Score 121; DB 1; Length 3175;
 Best Local Similarity 21.9%; Pred. No. 0.021;
 Matches 63; Conservative 37; Mismatches 94; Indels 94; Gaps 11;
 QY 3 LELDCGNSFKRWIVHADVIEGGIVDSQA-----LVAEVAALASVRLTGC 52
 DB 91 LELQHPAQLRVVDARLAIEASVISTDHASAKRFGARFALTPYANAWVSPAAN 150
 QY 53 IVSVRSPEETDLCALIAQAFVAQKVAHPVREMAGVNGYDDVQ-----RLGMDRW 104
 DB 151 SLIVTTQEQDQFCWL-----KLLPDRREGAGLRLYNNHREORTGWLSTGLRLW 201
 QY 105 LAALG-----AFHLAKAG-----LVIDLGTAAKADFVSADGEH 138
 DB 202 LGDLGLGINASSGGLKPHIMRSPORAWHITRSCCLKSYVCDI---SEADRSCLPAGN 258
 QY 139 LGGYICGPMRLMSQLRTHTRRYDDASAEALSSLSPGRSVTVEAVERGVLMQLGFAY 198
 DB 259 YGGYNPPG-----DGACGYRCLAFMN-GATVVSA---GC-----288
 QY 199 TQLEQARVLAGEEFTVLTGDAPLVRAALPQARVVPDLVFGVLAAC 246
 DB 289 -----SSDLWCDDDELATRVQLSFTFTVTIPGGRVCPNARY----AMIC 328

RESULT 4

SRPR_HUMAN STANDARD; PRT; 638 AA.
 ID SRPR_HUMAN STANDARD; PRT; 638 AA.
 AC P08240; Q9BVU4;
 DT 01-AUG-1998 (Rel. 08, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Signal recognition particle receptor alpha subunit (SR-alpha)
 DE (Docking protein alpha) (DP-alpha).
 GN SRPR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88124220; PubMed=3340536;
 RA Hortsch M., Labelt S., Meyer D.I.;
 RT "Complete cDNA sequence coding for human docking protein.";
 RL Nucleic Acids Res. 16:361-362(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, Muscle, and Placenta;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS INTEGRAL MEMBRANE PROTEIN ENSURES, IN CONJUNCTION
 CC WITH SRP, THE CORRECT TARGETING OF THE NASCENT SECRETORY PROTEINS
 CC TO THE ENDOPLASMIC RETICULUM MEMBRANE SYSTEM.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM MEMBRANE. THOUGHT TO
 CC BE ANCHORED IN THE MEMBRANE THROUGH AN INTERACTION WITH SR-BETA,
 CC WHICH CONTAINS A BONA FIDE TRANSMEMBRANE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X06272; CAA29608.1; --
 CC EMBL; BC001162; AAH01162.1; --
 CC EMBL; BC009110; AAH09110.1; --
 CC EMBL; BC013583; AAH13583.1; --
 CC PIR; A29440; A29440.
 CC HSSP; O07347; IFFH.
 CC Genew; HGNC:11307; SRPR.
 CC MIW; 182180; --
 CC InterPro: IPR000897; SRP54.
 CC Pfam; PF00448; SRP54; 1.
 CC Pfam; PF02881; SRP54_N; 1.
 CC ProDom; PD000819; SRP54; 1.
 CC PROSITE; PS00300; SRP54; 1.
 CC Signal recognition particle; Transmembrane; Receptor;
 CC Endoplasmic reticulum; Gtp-binding.
 CC TRANSMEM 11 33 POTENTIAL.
 CC NP_BIND 425 432 GTP (BY SIMILARITY).
 CC NP_BIND 520 524 GTP (BY SIMILARITY).
 CC NP_BIND 588 591 GTP (BY SIMILARITY).
 CC CONFLICT 81 81 K -> R (IN REF. 1).
 CC CONFLICT 201 201 G -> E (IN REF. 1).
 CC CONFLICT 306 307 NS -> TL (IN REF. 1).
 CC SEQUENCE 638 AA; 69811 MW; 967F943CEE3FA79E CRC64;

Query Match 7.7%; Score 98; DB 1; Length 638;
 Best Local Similarity 25.9%; Pred. No. 0.39;
 Matches 63; Conservative 30; Mismatches 84; Indels 66; Gaps 12;
 QY 33 DOALVAEVAALASVRLTGCRTVSVRSEET-----DALCALTAQAFVQ-AKVAHPVRMA 87
 DB 347 DHLIAKNVAADIADVQ--CESVANKLEGKVGMTSTVTSTVKQALQESLVQILQPRRD 404

CC	EMBL; D26185; BAA05305.1; -;
DR	EMBL; Z99104; CAB11846.1; -;
DR	Subtilist; EG10133; yacB.
DR	InterPro; IPR004619; Baf.
DR	Pfam; PF03309; Bvg_acc_factor; 1.
DR	TIGRFAMs; TIGR00671; baf; 1.
KW	Hypothetical protein; Complete proteome.
SO	SEQUENCE 233 AA: 26217 MW: 8AE96E732C15DF44 CRC64:

Query Match	10.28;	Score 129.5;	DB 1;	Length 233;
Best Local Similarity	24.89;	Pred. No. 0.00022;		
Matches	38;	Conservative 36;	Mismatches 56;	Indels 23; Gaps 7;
Qy	64	ALCALIAQAFVAQAKVAHPVREMGVRNG----	YDDYQRLGMDRWLAALAGAFHLKAGACL	119
Db	72	ALERMCKYTHIEPQIVGP-----	GMTGLNLIKVDNPEKVGADRIYVAVAAIHLYGNPLI	126
Qy	120	VIDLGTAAKADFVSADGEHLGGYICPGMPLRMSOLRTHTRRI--	RYDDASAEFRALSSLSPG	178
Db	127	VWDFGTATTTCYIDENQYMGGAAPGITTISTEALYSRAAKLPRIETRPDNI	-----G	181
Qy	179	RSTVEAVERGCVLMLQGFAYT-QLE--	QARVLW	208
Db	182	KNTVSAQMOSGIL-----	FGYVGQVEGIVKRMKW	209

RESULT 3	
RPOA_EAV	
ID	RPOA_EAV STANDARD; PRT; 3175 AA.
AC	P19811; Q88625;
AD	01-FEB-1991 (Rel. 17, Created)
DT	01-DEC-1992 (Rel. 24, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	POL polyprotein (ORF1a b) [Contains: RNA-directed RNA polymerase
DE	(EC 2.7.7.48); Helicase; Protease (EC 3.4.21.-)].
GN	POL.
OS	Equine arteritis virus (EAV).
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC	Arteriviridae; Arterivirus.
OX	NCBI_TaxID=11047;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Bucyrus;
RX	MEDLINE=91237805; PubMed=1851863;
RA	den Boen J.A., Snijder E.J., Chirnside E.D., de Vries A.A.F.,
RA	Horzinek M.C., Spaan W.J.M.;
RT	"Equine arteritis virus is not a togavirus but belongs to the
FT	coronavirusslike superfamily".
RL	J. Virol. 65:2910-2920(1991).
RN	[2]
RP	SEQUENCE OF 1-17 FROM N.A.
RC	STRAIN=Bucyrus;
RX	MEDLINE=30287699; PubMed=2162519;
RA	de Vries A.A.F., Chirnside E.D., Bredjenbeek P.J., Gravestien L.A.,
RA	Horzinek M.C., Spaan W.J.M.;

RESULT 12

PCT-US94-06362-4
; Sequence 4, Application PC/TUS9406362
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: OHTA, YASUHIRO
; TITLE OF INVENTION: Stress Proteins and Uses Therefor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06362
; FILING DATE: 08-JUN-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,381
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI88-08AFA2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-06362-4

Query Match 6.9%; Score 87.5; DB 5; Length 540;
Best Local Similarity 22.4%; Pred. No. 0.087;
Matches 58; Conservative 27; Mismatches 79; Indels 95; Gaps 11;

QY 22 AVIEGGIVDSQALVAEVAALA-----SVRLTGCRIVSVSEETDLCALIAQAFV 75
Db 291 AILTGQVISEEVLLENADLSLGLKARKVVTDETTIVEGAGTDAIGRVAQ----- 346
76 QAKVAHPVREMAGVRNGYDDYQRLGMDRWLAALGAFHLAKGACLVLDLGTAAKADFVSAD 135
Db 347 -----IRQ--EIESDSYDREKLOERLAKLA-----GGVAVLKAGAA----- 382
QY 136 GEHLGVCYCPMLPSQLRTHRRYDDDAEALSSLSRSTVEAVERGCVLMLOG 195
Db 383 -----TEVELKERKHIEDAV-----RNAKAAVEEGIV-----AG 412
QY 196 FAYTOLEQARV-----LWGEFTF---VELTGGDAPLVRAA-----LP-- 229
Db 413 GGVTLQQAPTLDLDELGEDEATGANIVKVALEAPLQIAFNAGLEPGVVAEKVRNLPAG 472
-QY 230 -----QARVVVDLVFVGLA 243
Db 473 HGLNAQTGYVEDLLAAGVA 491

- RESULT 13

US-08-553-703A-1
; Sequence 1, Application US/08553703A
; Patent No. 5795767
; GENERAL INFORMATION:
; APPLICANT: MARU, ISAFUMI

APPLICANT: OHTA, YASUHIRO
APPLICANT: TSUKADA, YOJI
TITLE OF INVENTION: EPIMERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,703A
FILING DATE: 30-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-553-703A-1

Query Match 6.7%; Score 85.5; DB 1; Length 402;
Best Local Similarity 24.3%; Pred. No. 0.096;
Matches 61; Conservative 25; Mismatches 78; Indels 87; Gaps 15;

QY 14 WRV-----IRVADAVI-----EGGIVDSQALVAEVAALASVRLTGCRIVSV 56
Db 135 WRVTAERYQSEAVDMQIVHVREDPSGLGRPQLPGAVASESWAVPMMLL--CLVEQL 192
QY 57 RSEETDLCALIAQAFVQAKVAHPVREMAGVRNGYDDYQRLGMDRWLAALGAFHLAKG 116
Db 193 GEED-----ELAG-----RYAQLG--HWCARRILQHVQRD 221
QY 117 ACVLVDLGTAAKADFVSADGEHLGVCYCPMLPSQLRTHRR-----IRYDDASAE 171
Db 222 GOAVL-----ENYSEDGEELSG--CLG-----RHQNPGHAEAGWFLRLHSSRSGDAK 267
QY 172 LSS-----LSPGRSTVEAVERGCVLMLO---GFAYTOLEQARVW---GEEFTVLTG 218
Db 268 LRAHVDTFLLPLFFRSGWDA-DHGGFLTFQDADGLCPTOLEWAMKLNPHSEAMTAFMG 326
QY 219 ----GDAPLVR 225
Db 327 YSESGDPALLR 337

RESULT 14

US-09-006-021-1
; Sequence 1, Application US/09006021
; Patent No. 5994105
; GENERAL INFORMATION:
; APPLICANT: MARU, ISAFUMI
; APPLICANT: OHTA, YASUHIRO

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/336,251
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06362
; FILING DATE: 06-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,381
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/804,632
; FILING DATE: 15-JUN-1989
; APPLICATION NUMBER: US 07/207,298
; FILING DATE: 09-DEC-1991
; APPLICATION NUMBER: US 07/366,581
; FILING DATE: 15-JUN-1988
; APPLICATION NUMBER: PCT/US89/02619
; FILING DATE: 15-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI88-08AFA4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-461-722-4

Query Match 6.9%; Score 87.5; DB 4; Length 540;
Best Local Similarity 22.4%; Pred. No. 0.087; 79; Indels 95; Gaps 11;
Matches 58; Conservative 27; Mismatches 27;

Qy 22 AVIEGGIVDSQALVAEVAALA-----SVRLTGCRIVSVEEETDALCALIAQAFV 75
Db 291 AILTGQGVISEEVLTLNADLSLLGKARKVVTKEITIVGAGDTDAIAGRVAQ---- 346
Qy 76 QAKVAHPVREMGVNGVDDYQRLGMDRWLAALGAFHLAKGACLVLDLGTAAKADFSAD 135
Db 347 -----IRQ-----EIENSDDYDREKLOERLAKLA-----GGVAIVKAGAA-- 382
Qy 136 GEHLGGYICPGMLRSLRTHRRIRYDDASAEALSSLSFGSTVEAVERGCVLMLOG 195
Db 393 -----TEVELKERKHIEDAV-----RNAAKAEVGV---AG 412
Qy 196 FAYTQLEQARY-----LWGEFT---VFLTGGDAPLVRAA-----LP--- 229
Db 413 GGVTLQAAPTLDLKLKLGDEATCANIVKVALEAPLQIAFNGLPGVVAEKVRLNLPAG 472
Qy 230 -----QARVVDLIVFVGLA 243
Db 473 HGLNAQTGVYEDLLAAGVA 491

RESULT 11
US-08-336-251-4
; Sequence 4, Application US/08336251
; Patent No. 6338952
; GENERAL INFORMATION:
; APPLICANT: Young, Richard S.
; TITLE OF INVENTION: Stress Proteins and Uses Therefor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,251
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06362
; FILING DATE: 06-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,381
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/804,632
; FILING DATE: 09-DEC-1991
; APPLICATION NUMBER: US 07/366,581
; FILING DATE: 15-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/207,298
; FILING DATE: 15-JUN-1988
; APPLICATION NUMBER: PCT/US89/02619
; FILING DATE: 15-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI88-08AFA3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-336-251-4

Query Match 6.9%; Score 87.5; DB 4; Length 540;
Best Local Similarity 22.4%; Pred. No. 0.087; 79; Indels 95; Gaps 11;
Matches 58; Conservative 27; Mismatches 27;

Qy 22 AVIEGGIVDSQALVAEVAALA-----SVRLTGCRIVSVEEETDALCALIAQAFV 75
Db 291 AILTGQGVISEEVLTLNADLSLLGKARKVVTKEITIVGAGDTDAIAGRVAQ---- 346
Qy 76 QAKVAHPVREMGVNGVDDYQRLGMDRWLAALGAFHLAKGACLVLDLGTAAKADFSAD 135
Db 347 -----IRQ-----EIENSDDYDREKLOERLAKLA-----GGVAIVKAGAA-- 382
Qy 136 GEHLGGYICPGMLRSLRTHRRIRYDDASAEALSSLSFGSTVEAVERGCVLMLOG 195
Db 393 -----TEVELKERKHIEDAV-----RNAAKAEVGV---AG 412
Qy 196 FAYTQLEQARY-----LWGEFT---VFLTGGDAPLVRAA-----LP--- 229
Db 413 GGVTLQAAPTLDLKLKLGDEATCANIVKVALEAPLQIAFNGLPGVVAEKVRLNLPAG 472
Qy 230 -----QARVVDLIVFVGLA 243
Db 473 HGLNAQTGVYEDLLAAGVA 491

```

Matches	71: Conservative	22: Mismatches	87: Indels	90: Gaps	14:
Qy	2	I L E D C G N S F T K R W I H V A D A V I E G G	---	I V D S D Q A L V A E V A A L S V R L T G C R I V S V R	57
Db	1134	L L Q R P S G E L W C H A V R N N H G H O T P D R Q G A D F W Y V D S S G A V V E C G L V A Q R L P G	---	---	1189
Qy	58	S E E E T D A L C A L I A Q A F A V	---	Q A K V A H P V R E M A G V R N G Y D D Y Q R L G M D R W L	107
Db	1190	R E E D D W F L E W E P A A V C T A K V	---	---	1229
Qy	108	L G A F H L A G A C L I V D I G T A A K A D F V S A	---	D G E	150
Db	1230	L R A M L E A G H A V H V A A E N N T S A A G Y R A L L A K A F D G Q A P T A V V H L G S L D G G G E L D P G L G A Q	---	---	1289
Qy	151	R S Q L R T H T R R Y D D A	---	S A E R A L S S L P G R S T V E A V E R G C V L M L	205
Db	1290	G A L	---	D A P R S A D V S P A L D P	1332
Qy	206	V L W G E E F T V L T	---	G G D A P L V R A A L	228
Db	1333	R L W	---	L L T R G A Q A V G A G D V S Y T Q A P L	1356

RESULT 8

US-09-568-472-4

```

: Sequence 4, Application US/09568472
: Patent No. 6358719
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE 1
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/17
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,4
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: . LENGTH: 1832
: . TYPE: PRT
: ORGANISM: Sorangium cellulosum
US-09-568-472-4

```

Query Match 7.18; Score 90; DB 4; Length 1832;

RESULT 9

```

US-09-567-899-4
; Sequence 4, Application US/095567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE PRODUCTION OF
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/095567899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,414
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1832
; TYPE: FRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-4

```

Query Match		7.1%; Score 90; DB 4; Length 1832;
Best Local Similarity		26.3%; Pred. No. 0.29;
Matches	71; Conservative	22; Mismatches 87; Indels 90; Gaps 14;
QY	2 ILEDCGNSFIKRWVIHVADAVIEGGG----	IVDSDQALVAEYAAASVRLTCRIVSVR 57
	: :::	: :::
Db	1134 LLSPSGELWCARVVNHGHQTDRQGADEFWVVDSSGAVVAEVCGLVAQRLLPG----	GVR 1189
	: :::	: :::
QY	58 SEETDLCALLTAQAFAV-OAKVAHPVREMAVGNYDDYQRLGMDRWL-----	AA 107
	: :::	: :::
Db	1190 REEDDFTELEWEPAAVGTAKV-----	NAGRWLILGGGGIGLA 1229
	: :::	: :::
QY	108 LGAFHLKAGACLVIDLGTAAKADFVSA-----DGE-----	H-----GGYICPGMPLM 150
	: :::	: :::
Db	1230 LRAMLEAGGHAVVHAAENNTSAAGVRIALLAKAPDGQAPTAVVHLSGLDGGELDPLDGLGAQ	1289
	: :::	: :::
QY	151 RSOLRTHTRIRYDDA--SARERALSSLPGRSTVEAVERGCVLML---	OGFAYTQLEQAR 205
	: :::	: :::
Db	1290 GAL-----DAPRGDVSPDALDP-----	ALVRGCDSVLTWVQALMGAGFDAP 1332
	: :::	: :::
QY	206 VLWGEEFTVFLT-----	GGDAFLVRAAL 228
	: :::	: :::
Db	1333 RLW-----LLTRGAOAGVGDVSVTQAPL	1356
	: :::	: :::

RESULT 10

```

US-08-461-722-4
; Sequence 4, Application US/08461722
; Patent No. 6335183
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A. and Young, Douglas
; TITLE OF INVENTION: Stress Proteins and Uses Therefor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,722
; FILING DATE: 05-JUNE-1995

```

Db 1134 LLQPSGELWCHARVNNHGHOTPDROGADFWVVDSSGAVVAEVCGLVAQRLPG-----GVR 1189
QY 58 SEETDLCALIAQAFV--QAKVAHPVREMAGVRNGYDYORLGMDRWL-----AA 107
Db 1190 REEDDWFLEWEPAAVGTAKV-----NAGRWLLGGGGGLGAA 1229
QY 108 LGAFHLKAGACLVLDLGTAAKADFVSA-----DGE-----HL-----GGYICPGMPLM 150
Db 1230 LRAMLEAGGHAVVHAENNTSAAGVRLAKAFDQAPTAVVHLGSLDGGGELDPGLGAQ 1289
QY 151 RSLQRTTRIRYDDA--SAERALSSLSPGRSTVEAVERGCVLML---OGFAYTQLEQAR 205
Db 1290 GAL-----DAPRSADVSPDALDP-----ALVRGCDSVLMTVQALAGMGRDAP 1332
QY 206 VLWGEFTVFLT-----GGDAPLVRAL 228
Db 1333 RLW-----LLTRGAQAVGAGDVSVTQAPL 1356

RESULT 5

US-09-567-969-4

; Sequence 4, Application US/09567969
; Patent No. 6355457

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/567,969

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1832

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-567-969-4

Query Match 7.1%; Score 90; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 0.29;
Matches 71; Conservative 22; Mismatches 87; Indels 90; Gaps 14;

QY 2 ILEDCGNSFIKRWIVHADVIEGG-----IVDSQALVAEVAALASVRLTGCRIVSVR 57
Db 1134 LLQPSGELWCHARVNNHGHOTPDROGADFWVVDSSGAVVAEVCGLVAQRLPG---GVR 1189
QY 58 SEETDLCALIAQAFV--QAKVAHPVREMAGVRNGYDYORLGMDRWL-----AA 107
Db 1190 REEDDWFLEWEPAAVGTAKV-----NAGRWLLGGGGGLGAA 1229
QY 108 LGAFHLKAGACLVLDLGTAAKADFVSA-----DGE-----HL-----GGYICPGMPLM 150
Db 1230 LRAMLEAGGHAVVHAENNTSAAGVRLAKAFDQAPTAVVHLGSLDGGGELDPGLGAQ 1289
QY 151 RSLQRTTRIRYDDA--SAERALSSLSPGRSTVEAVERGCVLML---OGFAYTQLEQAR 205
Db 1290 GAL-----DAPRSADVSPDALDP-----ALVRGCDSVLMTVQALAGMGRDAP 1332
QY 206 VLWGEFTVFLT-----GGDAPLVRAL 228
Db 1333 RLW-----LLTRGAQAVGAGDVSVTQAPL 1356

RESULT 6

US-09-568-480-4

; Sequence 4, Application US/09568480
; Patent No. 6355458

; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1832
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-4

Query Match 7.1%; Score 90; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 0.29;
Matches 71; Conservative 22; Mismatches 87; Indels 90; Gaps 14;

QY 2 ILEDCGNSFIKRWIVHADVIEGG-----IVDSQALVAEVAALASVRLTGCRIVSVR 57
Db 1134 LLQPSGELWCHARVNNHGHOTPDROGADFWVVDSSGAVVAEVCGLVAQRLPG---GVR 1189
QY 58 SEETDLCALIAQAFV--QAKVAHPVREMAGVRNGYDYORLGMDRWL-----AA 107
Db 1190 REEDDWFLEWEPAAVGTAKV-----NAGRWLLGGGGGLGAA 1229
QY 108 LGAFHLKAGACLVLDLGTAAKADFVSA-----DGE-----HL-----GGYICPGMPLM 150
Db 1230 LRAMLEAGGHAVVHAENNTSAAGVRLAKAFDQAPTAVVHLGSLDGGGELDPGLGAQ 1289
QY 151 RSLQRTTRIRYDDA--SAERALSSLSPGRSTVEAVERGCVLML---OGFAYTQLEQAR 205
Db 1290 GAL-----DAPRSADVSPDALDP-----ALVRGCDSVLMTVQALAGMGRDAP 1332
QY 206 VLWGEFTVFLT-----GGDAPLVRAL 228
Db 1333 RLW-----LLTRGAQAVGAGDVSVTQAPL 1356

RESULT 7

US-09-568-486-4

; Sequence 4, Application US/09568486
; Patent No. 6355459

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,486

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1832

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-568-486-4

Query Match 7.1%; Score 90; DB 4; Length 1832;
Best Local Similarity 26.3%; Pred. No. 0.29;

Db 174 LGQQTGQLPQVEM---EAIKSLPPRFALNTTEAIOSGVI-----YTLIAGMRDFTTEW 223
QY 212 FTVF-----LTGGDAPL-----VRAALPO--ARVV--PDLVFGVL 242
Db 224 LSLFPGKVAIKGGDRILLNLYLOALYPDLAARLIVEPNLIIFWGM 268

Search completed: June 24, 2003, 22:16:21
Job time : 22.9032 secs

RA	Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA	Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA	Trust T.J.;
RT	"Genomic sequence comparison of two unrelated isolates of the human
RT	gastric pathogen Helicobacter pylori.";
RL	Nature 397:176-180(1999).
DR	EMBL; AE001509; AAD06372.1; -.
DR	InterPro; IPRO04619; Baf.
DR	Pfam; PF03309; Bvg_acc_factor; 1.
DR	TIGREMS; TIGR00671; baf; 1.
KW	Complete proteome.
SQ	SEQUENCE 223 AA; 24754 MW; B31377DA1AEF6E0E CRC64;
Query Match	10.8%; Score 137.5; DB 16; Length 223;
Best Local Similarity	27.8%; Pred. No. 0.00068;
Matches	57; Conservative 30; Mismatches 63; Indels 55; Gaps 10;
QY	54 VSVSEETDAL-CALIAQAFVQAIVP-VREMAGVRNGDDYQRLGMDRWLAALGAF 111
DB	ISVNEENKALLNC-----YPNAKNIAAGFFHLETDYIGLGIDROMACLAVV 96
QY	112 HLAGKACGLVDLGTAAKADFVSADGHELGSGYCPCGMPLMRSQLRTHTRIRYDDASAE-- 169
DB	GVIVDAGSATIDLK-EGRHLCGLDPGL-----AQVHYAKYSKAIL 140
QY	170 ----RALSSLSP--GRSTVEAVERGCVL----MLQGFPAYTQLEQARVLWGEEFTVFLTGG 220
DB	EQPFRALDSLEVLPKPTRDAVNVMILSIICQHAKDQ-----KIYLCCGD 188
QY	221 ADLVRAALPOARVPVDPDLVFGVGLAMA 245
DB	AKYLSAFLPHSVCKERLVEDGMEIA 213
RESULT 15	
ID	Q8YQD7 PRELIMINARY; PRT; 276 AA.
QY	Q8YQD7;
DT	01-WAR-2002 (TrEMBLrel. 20, Created)
DT	01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Hypothetical protein Alr3896.
GN	Alr3896.
OS	Anabaena sp. (strain PCC 7120).
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
NCBI_TaxID=103690;	
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21595285; PubMed=11759840;
RA	Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA	Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA	Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA	Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA	Yasuda M., Tabata S.;
RT	"Complete genomic sequence of the filamentous nitrogen-fixing
RT	cyanobacterium Anabaena sp. strain PCC 7120.";
RL	DNA Res. 8:205-213(2001).
DR	EMBL; AP003594; BAB75595.1; -.
DR	InterPro; IPRO04619; Baf.
DR	Pfam; PF03309; Bvg_acc_factor; 1.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 276 AA; 30272 MW; 2AD04CA693E56F25 CRC64;
Query Match	10.0%; Score 127; DB 16; Length 276;
Best Local Similarity	32.1%; Pred. No. 0.0068;
Matches	53; Conservative 24; Mismatches 58; Indels 30; Gaps 9;
QY	96 YQRLGMDRWLAALGAFLHAKGACGLVLDLGTAAKADFVSADG--EHLGGVICPGMPLMRSQ 153
DB	YPTLGLIDRALALWAGMSGWGFVPLVIDAGTA--LFETAADGKNLVGGAILFGVGLQFPAS 173
QY	154 LRTHTRIRYDDASAERALSLSLP--GRSTVEAVERGCVLMLQGFAYTQLEQARVLWGEE 211

RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AE005867; AAK23910.1; --
 DR TIGR: CC1935; --
 DR InterPro: IPR004619; Baf.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMS: TIGR00671; baf; 1.
 DR Complete proteome.
 SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 12.1%; Score 153.5; DB 16; Length 261;
 Best Local Similarity 27.4%; Pred. No. 3.7e-05;
 Matches 66; Conservative 39; Mismatches 113; Indels 23; Gaps 8;
 QY 1 MILEDCGNSFIKRWIRVHVADAVIEGGGIVDSQA---LVAEVAALASVRLTGCR-----52
 Db 2 MLLAIEQNTNTMFAIHDCASVQWRSATSTRADYVWLSQLLSQGLGFRDAIDAV 61
 QY 53 IVSVRSEETDLCALIAQAFVQAKVAHPVREMA--GVRNGYDDYQRLGMDRWLAALGA 110
 Db 62 IISVVPOSIFNLNLSRYFNVEPLV---IGENAKLGIDVRIEKPSEAGADRLVNAIGA 118
 111 FHLKAGACLVLDGTAAKADFSADGEHLGGVYICPGMPLMRSLRTHTRIRYDDASAE 170
 Db 119 AMVYGPVLVDSGTATTDIIVADGAFEGGIIAPGINSQAL--HEAAKLPRIATIOR 176
 QY 171 ALSSLSPGRSTVEAVERGC-----VLMQGFAYTQLEQARVLMGEFTVLTGGDAPLVRA 226
 Db 177 PAGNRIVGDTVYSMQSGVFWGVISLIEGLV-ARIKAER--GEPMTVIATGVSALFPG 232
 QY 227 A 227
 Db 233 A 233

RESULT 11

Q8XHL5 PRELIMINARY; PRT; 259 AA.
 ID Q8XHL5;
 AC Q8XHL5;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein CPE2468.
 GN CPE2468.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1502;
 RN [1]
 SEQUENCE FROM N.A.
 STRAIN=13 / TYPE A;
 PubMed=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of *Clostridium perfringens*, an anaerobic
 RT flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL: AF003194; BAB82174.1; --
 DR InterPro: IPR004619; Baf.
 DR InterPro: IPR000515; BPD.transp.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMS: TIGR00671; baf; 1.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 259 AA; 28819 MW; OD5FEA3B7A145E10 CRC64;

Query Match 11.9%; Score 150.5; DB 16; Length 259;
 Best Local Similarity 24.5%; Pred. No. 6.5e-05;
 Matches 64; Conservative 45; Mismatches 85; Indels 67; Gaps 11;
 QY 1 MILEDCGNSFIKRWIRVHVADAVIEGGGIVDSQAALVAEVAALASVRLTGCRIVSRSE 60
 Db 1 MILLIDVGNTI-----VL-----GIHDNEK-----YIASWR-----ISTDSKK 34

QY 61 ETDALCALIAQAF-----AVQAKVAHPVREMA-----GVRN 91
 Db 35 TSDEYSIQVQQLFNQAKLPEDVEGIIISVVPVNMHSLNVMVRKCFCKEPIVVGPIKT 94
 QY 92 G-----YDDYQRLGMDRWLAALGAFHLAKGACLVLDGTAAKADFSADGEHLGGVYICPGM 147
 Db 95 GINIKYDNPKEGADRIVNAVAFAEKHKPMIIDFGTATTTCATTEKGDYLGNGICPGI 154
 QY 148 PLMRSLRTHTRIRYDDASAEALSSLSPGRSTVEAVERGCVLMQGLQ---FAYTQLEQA 204
 Db 155 QISADAL--FERAAKLPRIELEKPSVIC--RNTVTSMQAGIIVYIGVKVEYIVRNMKKE 210
 QY 205 RVLWGE-EFTVLTGGDAPLV 224
 Db 211 MMDLGEKEFPVLATGGLAKLV 231
 RESULT 12
 Q8R7M2 PRELIMINARY; PRT; 255 AA.
 ID Q8R7M2;
 AC Q8R7M2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative transcriptional regulator, homologs of Bvg accessory
 DE factor.
 GN TTE2381.
 OS Thermococcus tengcongensis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 SEQUENCE FROM N.A.
 STRAIN=MB4T / JCM11007;
 RC MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of *T. tengcongensis* genome.";
 RL Genome Res. 12:689-700(2002).
 DR EMBL: AE013180; AAM25520.1; --
 DR Complete proteome; Hypothetical protein.
 SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;

Query Match 11.6%; Score 147.5; DB 16; Length 255;
 Best Local Similarity 22.8%; Pred. No. 0.00011;
 Matches 55; Conservative 45; Mismatches 110; Indels 31; Gaps 5;
 QY 1 MILEDCGNS-----FIKRWIRVHVADAVIEGGGIVDSQAALVAEVAALASVRLTGCRIV 54
 Db 1 MLLAFDVGNTNIVMGVFKGKLLHSFRISTDKNKTVDYEGMLVNQLIGYNGISLT-----55
 QY 55 SVRSEETDLCALIA-----QAFVQAKVAHPVREMAVGRNG-----YDDYQRLGMDR 103
 Db 56 -----EIDDIISVVVPLMNTLQVMSLRYFTKPIVVGPIKINIKYDNPKEVGADR 110
 QY 104 WLAALGAFHLAKGACLVLDGTAARADFSADGEHLGGVYICPGMPLMRSLRTHTRIRY 163
 Db 111 IVNAVAAYELGYPVIVDFGTATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 170
 QY 164 DDASAEALSSLSPGRSTVEAVERGCVLMQGLQFAYTQLEQARVLMGEFTVLTGGDAPL 223
 Db 171 IDLTPTPTVIN---RNTVASMOSGIIYGHVGMVDYIVTRMKGEFAPSYYVATGTFANN 226
 QY 224 V 224
 Db 227 I 227
 RESULT 13
 O25533 PRELIMINARY; PRT; 223 AA.
 ID O25533


```

Db 501 RYPPTTCNAVAS-----GMMDAV-CGSVMHMG-----RLKE-KTGAGKPDVDVITGGG 549
QY 221 APLVRAALPQA-----RVVPDLVFFVGL 242
Db 550 AAKVAEALPPAFLAENTVRVADNLVHGL 578

RESULT 4
QY2M4 PRELIMINARY; PRT; 295 AA.
AC Q8Y2M4;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative transcription regulation accessory factor transcription
DE regulator protein.
GN RSC0311 OR RSC03278.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_taxid=305;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646058; CAD13839.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR Complete proteome.
SQ SEQUENCE 295 AA; 30990 MW; 133074A7764BFCBB CRC64;

Query Match 18.8%; Score 239; DB 16; Length 295;
Best Local Similarity 33.6%; Pred. No. 2.5e-12;
Matches 92; Conservative 33; Mismatches 107; Indels 42; Gaps 14;

QY 2 ILEDCGNSFTK--WRVTHVA-DAVIEGG-----GIVDSQALVAEVA-----ALAS 45
Db 13 LLLLDAGNTRIKWATADVAPVAGGTPWQAGARPHDQ--LAELVEDWRDCHACAG 70
QY 46 VRLTGCRIVSRSEETDALCALIAQAF-AVQAKVAHPVREMAGVRNGYDDYQRLGMDRW 104
Db 71 MAPPDVWISVVAGPALRDALCARIARVFDGARLRIVASERAAAAGLRNGYRDPAQIGTDRW 130
QY 105 LAALGAFHL-AKGACLVIDIGTAAKADFSVSDGHEHGGYICPGMPLMRSQRLTRRI-- 161
Db 131 VGAVGARHAWPDTALLVTAGTATTLTLDIVAPDGRFAGGLILPGLTLMRLSRNLAQIPE 190
QY 162 -----RYDDASERALSLSLSPGRSTVEAVERGCVLMQGFAYTQLEQA-RVLWGEEFT 213
Db 191 IDIGYLARDDAQAPDVPSWAD--NTQDAIAGCVTAQAG-ALAQTWALQAOYPGPIR 247
QY 214 VFLTGGDAPLVAAL-POARVV-----PDLVFFVGL 242
Db 248 CVLSGG-----ARAALAPHLRMPFQMDNLVLLGL 277

RESULT 5
QYX8N6 PRELIMINARY; PRT; 265 AA.
AC QYX8N6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.

```

```

OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_taxid=1902;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3];
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4];
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL049628; CAB40880.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 16.5%; Score 210; DB 16; Length 265;
Best Local Similarity 28.6%; Pred. No. 6.2e-10;
Matches 81; Conservative 46; Mismatches 90; Indels 66; Gaps 14;

QY 1 MILEDCGNSFIKRWVTHVADAVIEGGIVD-----SDQALVAEVAALASVRLTGCRIVS 55
Db 1 MLTIDVGN-----HTVLGDFGDIVHEWRISTDSRRTADELAVLLQGLMGH--P 51
QY 56 VRSEETD-----ALCALIAQAFVQAKVAPVREMA-----GVRNGY--- 93
Db 52 LLGDELGDGIDGIAICATV-----PSVLHELREVTTRYGVDPVAVLPGVGTGPIL 104
QY 94 -DQYQRLGMDRWLAALGAFHLAKGACLVIDIGTAAKADFSVSDGHEHGGYICPGMPLMRS 152
Db 105 TDHPKEVGADRIINAAVAVELYGGPAIVVDFTATTEDAVSARGEYIGGVAPGIEISVE 164
QY 153 QLRTHTRIRYDDASAEERALSLSLSPGRSTVEAVERGCVLMQGFAYTQLE-----QARVL 207
Db 165 ALGVKGALRKRIEVARPSVI-----GNKTVEAMSGIV---YGFQ-GQVDGVVNRMAREL 216
QY 208 WG--EEFTVFLTGDDAPLVAALPQARVV-----PDLVFFVGLAM 244
Db 217 ADPPDDVTVIATGGLAPMV---LGESSVIDEHEPWLTMGLRL 256

RESULT 6
QYCD56 PRELIMINARY; PRT; 274 AA.
ID Q9CD56

```

```
Db 1 MLELDGNSLKWVIEGAARSV-AGGLAEDSDALVEOLTSQLPVRACRLSVSRSEQ 59
QY 61 ETDALCALIAQAFVAQVAHPVREMAGVNRGYDDYQRLGMDRWLAALGAFHLKAGACLV 120
Db 60 ETSQVLALEQLFPVSALVASSGKQAGVNRGYLDYQRLGMDRWLAALVNAHHLKAGACLV 119
QY 121 IDLGTAAKADFSADGEHLGGYICPGMPLMRSQLRTHRRIRYDDASAEALSSLSPGRS 180
Db 120 IDLGTAVTSDLVAAADGVHLGGYICPGMPLMRSQLRTHRRIRYDDAEARRALASLQPGA 179
QY 181 TVEAVERGCVLMQGFAYTQLEQARVLMGEETVFLTGGDAPLRAALPQARVVDLVFV 240
Db 180 TAEAVERGCLMLRGFVRREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPOLVFV 239
QY 241 GLAMACPLD 249
Db 240 GLALACPIE 248
RESULT 2
QJXF1 PRELIMINARY; PRT; 592 AA.
AC QJXF1
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE BIRA protein/Bvg accessory factor.
GN NMB2075.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
Gill J., Scarlato V., Masigiani V., Pizze G., Grandi G., Sun L.,
Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002557; AAF42394.1;
DR HSSP; P06709; 1BIA.
DR TIGR; NMB2075;
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BIRA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00121; bira_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 592 AA; 64701 MW; 9E27421DA2B41DE3 CRC64;
Query Match 19.3%; Score 245; DB 16; Length 592;
Best Local Similarity 30.1%; Pred. No. 1.9e-12;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
QY 3 LEIDCGNSIKRWVHVADAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
Db 341 LLLDGGNSRLKW-----AWVNGTFTVGSAPYRDLSPGLAEWAERKADGNVIVGC-- 391
QY 54 VSVRSEETDLCALIAQAFVAQVAHPV-----EMAGVNRGYDDYQRLGMDRWLA 107
Query Match 19.3%; Score 245; DB 16; Length 592;
Best Local Similarity 30.1%; Pred. No. 1.9e-12;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
```

```
Db 392 -----AVCGEFKA-QVQEQARKIEWLPSSAQALGIRNHYRHPPEHGSDFWNA 440
QY 108 LGAFHLKAGACLVLDLGTAAKADFSADGEHLGGYICPGMPLMRSQLRTHRRIRI----- 161
Db 441 LGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGTTIMPGFHLMKESLAVRTANLRHACK 500
QY 162 RYD-DASAEALSSLSPGRSVTEAVERGCVLMQGFAYTQLEQARVLMGEETVFLTGGD 220
Db 501 RYPPFTTTGNVAVS-----GMMDAV-CGSVMMHG-----RLKE-KTGACKPVDVITGGG 549
QY 221 APLVRAALPQA-----RVDPDLVFGV 242
Db 550 AAKVAEALPPAFLAENTVRVADNLVIYGL 578
RESULT 3
QJWI7 PRELIMINARY; PRT; 592 AA.
AC QJWI7
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE BIRA bifunctional protein (EC 6.3.4.15).
GN BIRA OR NMA0357.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
Whitehead S., Spratt B.G., Barrrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83659.1;
DR HSSP; P06709; 1BIA.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BIRA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00121; bira_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW Ligase; Complete proteome.
SQ SEQUENCE 592 AA; 64468 MW; 7ED2E0CD3B31C630 CRC64;
Query Match 19.1%; Score 243; DB 16; Length 592;
Best Local Similarity 30.1%; Pred. No. 2.8e-12;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
QY 3 LEIDCGNSIKRWVHVADAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
Db 341 LLLDGGNSRLKW-----AWVNGTFTVGSAPYRDLSPGLAEWAERKADGNVIVGC-- 391
QY 54 VSVRSEETDLCALIAQAFVAQVAHPV-----EMAGVNRGYDDYQRLGMDRWLA 107
Db 392 -----AVCGEFKA-QVQEQARKIEWLPSSAQALGIRNHYRHPPEHGSDFWNA 440
QY 108 LGAFHLKAGACLVLDLGTAAKADFSADGEHLGGYICPGMPLMRSQLRTHRRIRI----- 161
Db 441 LGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGTTIMPGFHLMKESLAVRTANLRHACK 500
QY 162 RYD-DASAEALSSLSPGRSVTEAVERGCVLMQGFAYTQLEQARVLMGEETVFLTGGD 220
Query Match 19.1%; Score 243; DB 16; Length 592;
Best Local Similarity 30.1%; Pred. No. 2.8e-12;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 21.9032 Seconds
(without alignments)
2342.388 Million cell updates/sec

Title: US-09-813-453A-61
Perfect score: 1269
Sequence: 1 MILEDCGNSFIKRWIRHVA.....QARWPDVFLVGLAMACPLD 249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	794.5	62.6	248	16 Q9HWC1	Q9hwc1 pseudomonas
2	245	19.3	592	16 Q9JXF1	Q9jxf1 neisseria m
3	243	19.1	592	16 Q9JW17	Q9jw17 neisseria m
4	239	18.8	295	16 Q8Y2M4	Q8y2m4 raiistonis s
5	210	16.5	265	16 Q9X8N6	Q9x8n6 streptomyce
6	176.5	13.9	274	16 Q9CD56	Q9cd56 mycobacteri
7	168.5	13.3	242	16 Q9PC14	Q9pci4 xylella fas
8	160.5	12.6	272	16 Q06282	Q06282 mycobacteri
9	159	12.5	273	16 Q97EB4	Q97eb4 clostridium
10	153.5	12.1	261	16 Q9A6Z1	Q9a6z1 caulobacter
11	150.5	11.9	259	16 Q8XHL5	Q8xhl5 clostridium
12	147.5	11.6	255	16 Q8R7M2	Q8r7m2 thermoanaer
13	144.5	11.4	223	16 Q25533	Q25533 helicobacte
14	137.5	10.8	223	16 Q9ZKY6	Q9zky6 helicobacte
15	127	10.0	276	16 Q8YQD7	Q8yqd7 anabaena sp
16	124	9.8	258	2 Q9F985	Q9f985 bacillus st

17	121	9.5	1727	12	Q8QZQ5	Q8qzq5 equine arte
18	121	9.5	3175	12	Q91DM2	Q91dm2 equine arte
19	118	9.3	257	16	P74045	P74045 synechocyst
20	118	9.3	273	16	O83446	O83446 treponema p
21	117.5	9.3	246	16	Q9WZY5	Q9wzy5 thermotoga
22	114.5	9.0	209	16	Q9PIA9	Q9pia9 campylobact
23	114.5	9.0	212	2	O32514	O32514 desulfovibr
24	114	9.0	254	16	Q9KGH5	Q9kg5 bacillus ha
25	109	8.6	259	16	O8YAC5	O8yac5 listeria mo
26	105	8.3	259	16	Q92F54	Q92f54 listeria in
27	98	7.7	576	4	Q96H02	Q96hu2 homo sapien
28	97.5	7.7	517	16	Q9RZD8	Q9rzd8 deinococcus
29	96.5	7.6	359	11	Q921H1	Q921h1 mus musculu
30	96.5	7.6	617	10	Q9LWU9	Q9lw9 oryza sativ
31	96.5	7.6	636	11	Q9DSG7	Q9ds7 mus musculu
32	96.5	7.6	636	11	O8VC45	O8vc45 mus musculu
33	96.5	7.6	940	16	Q9A6P6	Q9a6p6 caulobacter
34	95	7.5	262	16	Q9RX54	Q9rx54 deinococcus
35	94.5	7.4	561	16	Q9X857	Q9x857 streptomyce
36	92.5	7.3	229	16	O67753	O67753 aquifex aeo
37	91	7.2	354	17	O8ZYT5	O8zyt5 pyrobaculum
38	91	7.2	511	16	Q9L214	Q9l214 streptomyce
39	91	7.2	629	2	Q07639	Q07639 streptomyce
40	90	7.1	293	16	Q98JM9	Q98jm9 rhizobium 1
41	90	7.1	1832	2	Q9L8C8	Q9l8c8 polyangium
42	89	7.0	396	17	Q9HS71	Q9hs71 halobacteri
43	88.5	7.0	326	2	Q9RU0	Q9rbu0 burkholderi
44	88	6.9	1832	2	Q9K128	Q9k128 polyangium
45	87.5	6.9	332	2	Q8RPL4	Q8rpl4 anaplasma p

ALIGNMENTS

RESULT 1

Q9HWC1	ID	Q9HWC1	PRELIMINARY;	PRT;	248 AA.
AC	Q9HWC1:				
DT	01-MAR-2001	(TRENBLrel. 16, Created)			
DT	01-MAR-2001	(TRENBLrel. 16, Last sequence update)			
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)			
DE	Hypothetical protein PA4279.				
GN	PA4279.				
OS	Pseudomonas aeruginosa.				
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;				
OC	Pseudomonas.				
OX	NCBI_TaxID=287;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 15692 / PA01;				
RX	MEDLINE=20437337; PubMed=10984043;				
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,				
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,				
RA	Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,				
RA	Brody K.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,				
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,				
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;				
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an				
RT	opportunistic pathogen."				
RL	Nature 406:959-964(2000).				
DR	EMBL; AE004843; AAG07667.1; -				
DR	InterPro; IPR004619; Baf.				
DR	InterPro; IPR001230; Prenyl_site.				
DR	Pfam; PF03309; Bvg_acc_factor; 1.				
DR	PROSITE; PS00294; PRENYLATION; UNKNOWN_1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 248 AA; 26758 MW; 609D37480899DA3B CRC64;				

Query Match 62.6%; Score 794.5; DB 16; Length 248;
Best Local Similarity 63.5%; Pred. No. 2.3e-59;
Matches 158; Conservative 31; Mismatches 59; Indels 1; Gaps 1;
QY 1 MILEDCGNSFIKRWIRHVAADVIEGGIGVDSQALVAEVAALASVRLTGCRIVSRSEE 60

Query Match	12.4%	Score 157.5;	DB 9;	Length 272;
Best Local Similarity	23.6%;	Pred. No. 1.5e-07;		
Matches	59;	Conservative 36;	Mismatches 90;	Indels 65; Gaps 18;

QY	13	KWRV-----THVADAVIEGGGIVDSQALVAEVAALASVRLTGCRIVSVRSEETDALCA	67
	:	:	
	:	:	
	:	:	
Db	29	QWRITSEVTADELALTIDLGIED-----SERLTG-----	60
QY	68	LIAQAFVAQKVAHPVREM-----AGVRNG-----YDDYQRLGMDRWLAALG	109
	:	:	
	:	:	
Db	61	--TAALSTVPVLHEVRIMLDQWPSPVPHVIEPGVGTGIPLLVDNPKVEGDATVNCIA	118
QY	110	AFHLKAGACLVIDIGTAAKADFYSADGEHLGGVTCGPMPLMRSQRLTHRRIRYVDASAE	169
	:	:	
	:	:	
	:	:	
Db	119	AYDRFRKAAIIVDFGSSICVDVVSAKGEFLGGAIAEPGVQVSSDAAAARSAALURRVELARP	178
QY	170	RALSLSLSPGRSTVEAVERGCVLMLQGFAYVTLQEQAR-----VLWGEETVFVLTGGDAPLV	224
	:	:	
	:	:	
Db	179	RSV-----VGKNTVECMQAGAVFGFAGVLGDLVGRITEDYSGFSVDHDAIVATGHTAPLL	234
QY	225	RAALPQARVV	234
	:	:	
Db	235	---LPELHTV	241

```

RESULT 14
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Paterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 3
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

```

Query Match	12.3%	Score 156;	DB 9;	Length 250;
Best Local Similarity	22.6%	Pred. No. 1.8e-07;		
Matches	60;	Conservative 49;	Mismatches 81;	Gaps 10;
QY	1	MILEDGCGNSFIKRWVTHVDADVIEGGIVDSQAALVAE----	VAAALASVRLTGGRIVSV	56
Db	18	VILVDVGNT-----NIVLG---IYNDTKLTAEWRLSTDLRSAD EYGIQVMNL	63	
QY	57	RSEETD-----ALCALIAQAFVAQVAHPVREMAGVENG----	YD	94
Db	64	FQQDKLDPTLVGVIISSVVPNIMYSLEHMIRYFKINPLVGP-----	GIKTGNIKYD	118
QY	95	DYQRGLMDRMLAALGAPHLAKAGACLVLDLCTAAKADFSADGEHGGYICPCMPLMRSOL	154	
Db	119	NPKVEGADRIVNAVAHEIVKRSLLIIDFGTATTTCFAVRNGDYLGGAICPGIKYSSRAL	178	
QY	155	RHTTRTRYDDASAERALSLSPPG----RSTVEAVERGCVLMOGFAYTTQLQEARVLWGE	210	
Db	179	-----FEKAAKLPVELLIKPAYAIKCKNTISSIQSIGIV-----	YYLRQVKYLEF	223
QY	211	-----EFTVLTGCDAPLV	224	
Db	224	LKENLPDGRRTRTSLVLATGTCLAKLI	249	

RESULT 12
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Vocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24

```

RESULT 13
US-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712.363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-276

```


Db 501 RYPPFTTGNASV-----GMDAV-CGSVMMHG-----RLKE-KTGAGKPDVDTITGGG 549
QY 221 APLVRAALPOA-----RVPPDLVFGVL 242
Db 550 AAKVAEALPPAPLAENTVRVADNLVIYGL 578

RESULT 5
US-09-813-453A-22
; Sequence 22, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-813-453A-22

Query Match 19.1%; Score 243; DB 9; Length 592;
Best Local Similarity 30.1%; Pred. No. 1.6e-15;
Matches 81; Conservative 39; Mismatches 89; Indels 60; Gaps 12;
QY 3 LEIDCGNSFIKRWVHVADAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
Db 341 LLLDGNRLKW-----AWVNGTFATVGSAPYRDLSPGAEWAERKVDGNVRIGVC-- 391
QY 54 VSVRSEETDLCALIAQAFVQAKVAHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107
Db 392 -----AVCGEKKA-QVQEOLARKIEWLPSSAQALGIRNHYRHPHEGSDRWENA 440
QY 108 LGAFHLKAGCLVLDLGTAAKADFSVADGEHLGGVCGMPLMRSQRLTHTRRI----- 161
Db 441 LGSRRFSRNACVWSCGTAVTVDALTDGHLGGTIFGFLMKESLAVRTANLNHRHAGK 500
QY 162 RYD-DASAEALSSLSGCRSTVEAVERGCVLMLOGFAYTQLEQARVLMGEEFTVFLTGDD 220
Db 501 RYPPFTTGNASV-----GMDAV-CGSVMMHG-----RLKE-KTGAGKPDVDTITGGG 549
QY 221 APLVRAALPOA-----RVPPDLVFGVL 242
Db 550 AAKVAEALPPAPLAENTVRVADNLVIHGL 578

RESULT 6
US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-813-453A-39
Query Match 18.3%; Score 232; DB 9; Length 460;
Best Local Similarity 29.0%; Pred. No. 1.4e-14;
Matches 78; Conservative 41; Mismatches 90; Indels 60; Gaps 12;
QY 3 LEIDCGNSFIKRWVHVADAVIEGGIVDSQALVAEVAAL-----ASVRLTGCR 53
Db 209 LLEGGNSRLKW-----AWVNGTFATVGSAPYRDLSPGAEWAERKADGNVRIGVC-- 259
QY 54 VSVRSEETDLCALIAQAFVQAKVAHPVR-----EMAGVRNGYDDYQRLGMDRWLAA 107
Db 260 -----AVCGEKKA-QVQEOLARKIEWLPSSAQALGIRNHYRHPHEGSDRWENA 308
QY 108 LGAFHLKAGCLVLDLGTAAKADFSVADGEHLGGVCGMPLMRSQRLTHTRRI----- 161
Db 309 LGSRRFSRNACVWSCGTAVTVDALTDGHLGGTIFGFLMKESLAVRTANLNHRHAGK 368
QY 162 RYD-DASAEALSSLSGCRSTVEAVERGCVLMLOGFAYTQLEQARVLMGEEFTVFLTGDD 220
Db 369 RYPPFTTGNASV-----GMDAV-CGSVMMHG-----RLKEKKA-GKPDVDTITGGG 417
QY 221 APLVRAALPOA-----RVPPDLVFGVL 242
Db 418 AAKVAEALPPAPLAENTVRVADNLVIHGL 446

RESULT 7
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

Query Match 16.5%; Score 210; DB 9; Length 265;
Best Local Similarity 28.6%; Pred. No. 1e-12;
Matches 81; Conservative 46; Mismatches 90; Indels 66; Gaps 14;
QY 1 MLELDCCNSFIKRWVHVADAVIEGGIVD-----SQALVAEVAALASVRLTGCRIVS 55
Db 1 MLLTIDVGT-----HTVLGLFDGEDIVHEWRISTDSRRRLADELAVLQGLMGHM--P 51
QY 56 VRSEETD-----ALCALIAQAFVQAKVAHPVRMA-----GVRNGY--- 93
Db 52 LLGELGDIAGIACATV-----PSVLHELREVTTRYGDVPAVLVEGKVTGPIL 104
QY 94 -DDYQRLGMDRWLAAALGAFHLKAGCLVLDLGTAAKADFSVADGEHLGGVCGMPLMRS 152
Db 105 TDHPKEVGADRIINAVAEVLGGPAIVVDGCTATTEDAVSARGEYIGGVIAPGTEISVE 164

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1269	100.0	249	9	US-09-813-453A-61	Sequence 61, Appl	
2	849	66.9	249	9	US-09-813-453A-70	Sequence 70, Appl	
3	794.5	62.6	248	9	US-09-813-453A-20	Sequence 20, Appl	
4	245	19.3	592	9	US-09-813-453A-43	Sequence 43, Appl	
5	243	19.1	592	9	US-09-813-453A-22	Sequence 22, Appl	
6	232	18.3	460	9	US-09-813-453A-39	Sequence 39, Appl	
7	210	16.5	265	9	US-09-813-453A-4	Sequence 4, Appl	
8	181	14.3	244	9	US-09-813-453A-41	Sequence 41, Appl	
9	168.5	13.3	242	9	US-09-813-453A-65	Sequence 65, Appl	
10	163	12.8	267	9	US-09-813-453A-15	Sequence 15, Appl	
11	160.5	12.6	272	9	US-09-813-453A-5	Sequence 5, Appl	
12	159	12.5	255	9	US-09-813-453A-7	Sequence 7, Appl	
13	157.5	12.4	272	9	US-09-712-363-276	Sequence 276, Appl	
14	156	12.3	250	9	US-09-813-453A-3	Sequence 3, Appl	
15	154	12.1	258	9	US-09-813-453A-2	Sequence 2, Appl	
16	153.5	12.1	260	9	US-09-813-453A-51	Sequence 51, Appl	
17	144.5	11.4	223	9	US-09-895-913A-74	Sequence 74, Appl	
18	144.5	11.4	223	9	US-09-813-453A-14	Sequence 14, Appl	
19	144.5	11.4	223	9	US-09-813-453A-67	Sequence 67, Appl	

Db 171 AKLPRTITQAPETAIGKNTVHALQSLVFGVAEMVDGLLRIRAEPLG-----EAVAVAT 225
QY 228 GHTAPLLPELHTVDHYDQHLTLOGLRLVFFERNLEVO 264
Db 226 GGFSTVVGICQEIYDYDETTLRLGLVELWASRSEVR 262

RESULT 15

AAU91150
ID AAU91150 standard; Protein; 250 AA.

XX AC AAU91150;

XX DT 05-JUN-2002 (first entry)

XX DE Clostridium acetobutylicum pantothenate kinase Coax.

XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX NW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX NW Clostridium acetobutylicum.

XX NW WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX PR 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX DR WPI; 2002-269358/31.

XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -

XX PS Claim 10; Page 68-69; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.

XX SQ Sequence 250 AA;

Query Match 28.5%; Score 389; DB 23; Length 250;

Best Local Similarity 34.2%; Pred. No. 6.2e-31;

Matches 81; Conservative 63; Mismatches 85; Indels 8; Gaps 3;

QY 1 MLLAIDVNTHTVVGLLSGMKHAKVQOQWRIRTESEVTADALTIDGLIGEDS---ER 57

Db 18 VILVDVGNNTVILGIYN---DTKLTAEWRLSTDVLRSADEYGIQVWNLFQQDKLDPTL 73

QY 58 LTGTAAALSTVPSVLEHVRIMLDQYWPSPVPHVLEPFGVTRTGPLLVDPNKEVGADRYNCL 117

Db 74 VEGVIISWVFNIMYSLEHMIRKYFKINP-LVVGPGIKTGINKYDNPKKEVGADRYNAV 132

QY 118 AAYDFRKAATVDFGSSICVDVVSAGFEFLGGATAPGVQVSSDAARSALRRVELAR 177

Db 133 AAHELYKRLLIIDFGTATTCVARENGDYLGGACIGIKVKSSEALFEKAALPRVELIK 192

QY 178 PRSVVGKNTVECMOAGAVFGFAGLVGLVGRIREDSVSDHVDVAIVATGHTAPLL 234

Db 193 PAYAICKNTIISQSGIVYRQLRVQKYLFEKLENLPDGRRTTSLVLTATGGLAKLI 249

Search completed: June 24, 2003, 21:46:13
Job time : 28.9257 secs

Db	173	IAESSQIIKSTVSSMQAGIFVGFYGOCEGIIAEMKKQSNASPY-----VVAATGGLARM	227
QY	235	LPELHTVDYHDQHLTQGLRLVFRN	260
Db	228	TEKSSAVDILDFFLLKGLLELYRN	253
<p> RESULT 12 AAU91153 ID: AAU91153 standard; Protein; 258 AA. XX AAU91153; AC XX DT XX DT 05-JUN-2002 (first entry) XX Rhodobacter capsulatus pantothenate kinase Coax. DE XX Pantothenate kinase; Coax; antibiotic; antimicrobial; KW pantothenate kinase modulator; coenzyme A; bactericidal compound. KW XX Rhodobacter capsulatus. OS XX PN W0200215601-A2. XX PD 28-FEB-2002. XX PF 24-AUG-2001; 2001WO-US26531. XX PR 24-AUG-2000; 2000US-227860P. PR 20-MAR-2001; 2001US-0813453. XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC. XX PI Yocum RR, Patterson TA; XX DR WPI; 2002-369358/31. XX PT Identifying potential antibiotic or antimicrobial agent, comprises PT contacting composition comprising pantothenate kinase (Coax) protein PT with test compound and identifying inhibitor of the Coax protein - XX XX Claim 10; Page 71-72; 128pp; English. XX CC The invention describes assays for identifying a (potential) antibiotic CC comprising contacting an assay composition comprising a pantothenate CC kinase (Coax) protein with a test compound, and determining the ability CC of the test compound to inhibit the activity of the Coax protein, an CC essential enzyme for the production of coenzyme A. Coax protein is a CC valuable target for identifying bactericidal compounds. Coax modulating CC agents can be used in an infectious animal model to determine the CC efficacy, toxicity, or side effects of treatment with such an agent. This CC is the amino acid sequence of a pantothenate kinase (Coax) protein CC described in the invention. XX SQ Sequence 258 AA; Query Match 31.6%; Score 431.5; DB 23; Length 258; Best Local Similarity 37.6%; Pred. No. 3.3e-35; Matches 100; Conservative 46; Mismatches 101; Indels 19; Gaps 4 </p>			
QY	1	MLLAIDVRNTHTVVGLLSCKMEHAKVVOQQRIRTESEVTADE-----LALTIDGLIGED	54
Db	1	MLLICDCGNNTVFSVWDG---TDFAATWRIATDHRTADEYFVWLNTLMQLKGLGRI	56
QY	55	SERLGTGTAALSTVPSVLHEVRIMLDQQWPSVPHVLIEPGVVRTGIPLLVDNPKVEGADRV	114
Db	57	SEAIIS----STAPRVFNLRVLCNRVFDCCRPYVVGKPGCELPVAPRVDPGTGTPGDRLV	112
QY	115	NCLAAAYDRERKAAIIVDFGSSICVDVYSKAGEFLGGAIAPGVQVSSDAAAARSALRRVE	174
Db	113	NTVAGYDRHGGDLIVDFDGTATTFDVPADGAYIGGVIAIPGVNLISLEALHMAAALPHVD	172
QY	175	LARPSRVGKNTVECMQAGAFVGFAGLVGDLGVIRGRIRDEDSGFSVDHDAIVATGHTAPLL	234

XX	Bacillus stearothermophilus.
OS	
XX	WO200216601-A2.
PN	
XX	28-FEB-2002.
PD	
XX	
PD	
XX	24-AUG-2001; 2001WO-US26531.
PF	
XX	
PR	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
XX	WPI; 2002-2693358/31.
DR	
XX	N-PSDB; ABK54193.
XX	

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (CoaX) protein with test compound and identifying inhibitor of the CoaX protein -

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (CoaX) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the CoaX protein, an essential enzyme for the production of coenzyme A. CoaX protein is a valuable target for identifying bactericidal compounds. CoaX modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (CoaX) protein described in the invention.

Sequence 258 AA:

Query Match	35.1%	Score	479.5;	DB	23;	Length	258;
Best Local Similarity	37.8%	Pred.	No. 4.7e-40;				
Matches	101;	Conservative	56;	Mismatches	93;	Indels	17;
						Gaps	5;
QY	1	MLLAIDVYNHTVTVGLLSG--MKEHAKVVQOQRIRTESEVTADELALITDGL---IGSDS	55				
Db	1	MFVLDVGNNTVLVYDGDGLKHH-----WRIETSRKTEDEYGMKIKALLNHVGLQF	54				
QY	56	ERLTGTAALSTVPSVLHVEVRIMLDQYWPSPVPHVLEPGVYRTGIPILLVDNPNKEVGADRVN	115				
Db	55	SDIRGIILISVVPPINFALERMCLKTHKIP-LVVGPGIKTGIDIKYDNPKEVGADRVN	113				
QY	116	CLAAAYDRFKAAIYVDFGSSICVDVVSAGEFTLGGAIAPGVQVSDAAAARSAAALRRVEL	175				
Db	114	AVAGIHLVSGPLIIVDFGATTTCYINEKHQYMGGAIPGIMISTEALFARAALPRIEI	173				
QY	176	APRPSVVGKNTVECMQAGAVFGAGVLGLVGRIRDEYSGFSDVHDHVAIVATGHTAPLLL	235				
Db	174	ARPDDIIIGKNTVSAMQAGILYGVGVQVEGIVSRMKA-----KSKPPKVIATGGLAPLIA	228				
QY	236	PELTVHDYDQHLITLQGLRLVFERNLE	262				
Db	229	SESDIIDVDPFLTLTGKLLYBKNT	255				

RESULT 10
AAU91170
ID AAU91170 standard; Protein; 262 AA.
XX
XX
AAU91170;
XX
XX
DT 05-JUN-2002 (first entry)
XX
XX
DE Pantothenate kinase (Coax) #8.
XX

KW	Pantothenate kinase; Coax; antibiotic; a
KW	pantothenate kinase modulator; coenzyme
XX	
XX	Bacillus anthracis.
XX	
PN	WO200216601-A2.
XX	
XX	28-FEB-2002.
PD	
XX	24-AUG-2001; 2001WO-US26531.
PF	
XX	
XX	24-AUG-2000; 2000US-227860P.
PR	20-MAR-2001; 2001US-0813453.
XX	
XX	(OMNI-) OMNIGENE BIOPRODUCTS INC.
PA	
XX	Yocum RR, Patterson TA;
PI	
XX	
XX	WPI; 2002-2693358/31.
DR	N-PSD3; ABK54191.
DR	

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

Claim 8; Page 98-99; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

xx	SQ	Sequence	262 AA;
		Query Match	34.5%; Score 471.5; DB 23; Length 262;
		Best Local Similarity	35.2%; Pred. No. 3.1e-39;
		Matches	94; Conservative 67; Mismatches 93; Indels 13; Gaps 4
Qy	1	MLLAIDYRNRHTVVVGLLSGMKEHAKVVOOQRITSESVTADLALTIDGLI---GEDSER	57
		: : : : : : : : : : : : : : :	
Dd	1	MIFVLVDGNTNAVLGVF---EEGELRQHRWMTDRHKTEDEYGLMVKQLLEHGSLP	56
Qy	58	LTFGAALSTVPVSILHEVRIMLDQWPSPVPHVLIIEPGVTGTPLLVDPNPKVGADRIVNCL	117
		: : : : : : : : : : : : : : :	
Dd	57	VKGIIVSSVPPIMFALERMCYEYFKIKP-LVVGFGIKTGLNIKYENPREVGADRIVAV	115
Qy	118	AAYDRFKAATVPDFGGISICVDVYSKAGEFLGGTAIPGVYSSDAAAARSALRVVELAR	177
		: : : : : : : : : : : : : : :	
Dd	116	AGIHLYGSPLIIVDFGATTTCYYINEEKHYMGGVITPGIMISAELYGRAKLPRITTK	175
Qy	178	PSRVVGNVTBECMQAGVFGFAGLVGLVGRIBEDVSGFSVDHDVAIVATGHTAPLLPPE	230
		: : : : : : : : : : : : : : :	
Dd	176	PSSVVGNTVSAMQSGIILYGVGQVEGIVKRMKEAK-----QEPKVATIGGLAKLISEE	230
Qy	238	LHTVDHYDQHULTLOGLRILVFERNLEVQ	264
		: : : : : : : : : : : : : :	
Dd	231	SNVIDVDDPFFLTLAGLYMYERNANIQ	257

```

RESULT 11
ABBA47661
ID      ABBA47661 standard; Protein; 259 AA.
XX      XX
XX      AC      ABBA47661;
XX      XX
XX      DT      05-FEB-2002 (first entry)
XX      YY

```

```
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX N-PSDB; ABK54168.
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein -
XX Claim 10; Page 67-68; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX Sequence 258 AA;
Query Match 36.2%; Score 494.5; DB 23; Length 258;
Best Local Similarity 38.8%; Pred. No. 1.4e-41;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;
QY 1 MLLAIDVRNTHVTVGLLSGMEKHAQVQOQWIRTESEVTADALATIDGLIGDS---ER 57
DB 1 LLLVIDGNTNTVLGVY---HDGKLEYHWRLETSRHKTEDFCGMLRSLDFHGLMFEQ 56
QY 58 LTGTAALSTVPSVLHVRIMLDQWPSVPHVLEPGVRTGIPLLVDNPKEVGADRIYNCL 117
DB 57 IDGIISVVPPIFALERMCTKVFHIEPQI-VGPGMKTGLNIRKYNPNKEVGADRIYNV 115
QY 118 AAYDRFRKAAIIVDFGSSICVDVYSKAGEFLGGAIAFGVQVSSDAAAARSALRRVELAR 177
DB 116 AAHLHGPNLIVDFGTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETTR 175
QY 178 PRSVGKNTVCEMAGAVFGAGLVGIRREDVSGFSVDHDAIVATGHTAPLAPLE 237
DB 176 PDNIIGNTVSAMSGILFVGVEGLVRKMK-----WQAKQLKIATGTGGLAPLIANE 230
QY 238 LHTVDHYDQHLTLQGLRLVFRN 260
DB 231 SDCIDIYDPPTLAKGLELIYERN 253
RESULT 8
AAU91171
ID AAU91171 standard; Protein; 254 AA.
XX AC AAU91171;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #9.
XX DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus halodurans.
```

```
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX N-PSDB; ABK54192.
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein -
XX Claim 10; Page 100; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX Sequence 254 AA;
Query Match 35.7%; Score 487.5; DB 23; Length 254;
Best Local Similarity 37.9%; Pred. No. 7.2e-41;
Matches 100; Conservative 64; Mismatches 87; Indels 13; Gaps 4;
QY 1 MLLAIDVRNTHVTVGLLSGMEKHAQVQOQWIRTESEVTADALATIDGLI---GEDSER 57
DB 1 MILVIDGNTNTVLGVY---QDETLVHWRLETSRKTEDEYAMTVRSDFHAGLQFOD 56
QY 58 LTGTAALSTVPSVLHVRIMLDQWPSVPHVLEPGVRTGIPLLVDNPKEVGADRIYNCL 117
DB 57 IDGIVISVVPPIFALERMCTKVFHIEPQI-VGPGMKTGLNIRKYNPNKEVGADRIYNV 115
QY 118 AAYDRFRKAAIIVDFGSSICVDVYSKAGEFLGGAIAFGVQVSSDAAAARSALRRVELAR 177
DB 116 AAELYGYPALIVDFGTATTTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETTR 175
QY 178 PRSVGKNTVCEMAGAVFGAGLVGIRREDVSGFSVDHDAIVATGHTAPLAPLE 237
DB 176 PKQVGTNTIDSMQSGIFGYGVQVGVKRMKA-----QAESEPKVIATGGLAKLIGTE 230
QY 238 LHTVDHYDQHLTLQGLRLVFRNL 261
DB 231 SETIDVIDSFLTKGLQLIYRKNV 254
RESULT 9
AAU91172
ID AAU91172 standard; Protein; 258 AA.
XX AC AAU91172;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #10.
XX DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
```


PI Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 XX Claim 10; Page 69-70; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 XX Sequence 265 AA;

Query Match 51.1%; Score 698.5; DB 23; Length 265;
 Best Local Similarity 51.5%; Pred. No. 3.7e-62;
 Matches 138; Conservative 53; Mismatches 64; Indels 13; Gaps 3;
 Qy 1 MLLAIDVRNTHVTGVLGSGMKEHAKVVOQWRIRTESEVTADALATIDGLIG-----E 53
 Db 1 MLLTIDVGNTHVTGLGFDG----EDIYHWRISTDSRRTADELAVLLQGLMGHPLLGDE 56
 Qy 54 DSERLTGTAAALSTVPSVLHEVRIMLDQWPSVPHVLTPEGVRTGIPLLVDNPKKEVGADRI 113
 Db 57 LGDGDIGAIACATVPSVLHEVRITRYGDPFALVPEGVKGTGVPILTDPKKEVGADRI 116
 Qy 114 VNCIAAYDRFKAAIIVDFGSSICVDVVSAGKEFLGGAIAFGVQVSSDAAAARSALRRV 173
 Db 117 INAAVAEVLGGPAIVVDFGTATTFDVAARGEYIGVIAPIGIEISVEALGVKGAQLRKI 176
 Qy 174 ELARPSVGNKNTVECHQAGAVFGAGLVGDLVGRIRREDVSGFSDHDAIVATGHTAPL 233
 Db 177 EVARPSVGNKNTVEAMQSGIIVYGFAGQVGVNRMARELA--DDPDVTVIATGGILAPM 234
 Qy 234 LLPELHVVDHYDQHLTLOGLRLVFERNL 261
 Db 235 VLGESSVIDEHEPWLILMGLRLVYERNV 262

SUIT 4
 ID AAU91175 standard; Protein; 256 AA.
 XX AC AAU91175;
 XX
 XX 05-JUN-2002 (first entry)
 XX Pantothenate kinase (Coax) #13.
 XX
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX Clostridium difficile.
 XX
 XX WO200216601-A2.
 XX
 XX 28-FEB-2002.
 XX
 XX 24-AUG-2001; 2001WO-US26531.
 XX
 XX 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;
 PI
 XX WPI; 2002-269358/31.
 DR N-PSDB; ABK54196.
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 XX Claim 6; Page 105; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 XX Sequence 256 AA;

Query Match 37.5%; Score 513; DB 23; Length 256;
 Best Local Similarity 38.9%; Pred. No. 1.9e-43;
 Matches 103; Conservative 65; Mismatches 85; Indels 12; Gaps 4;
 Qy 1 MLLAIDVRNTHVTGVLGSGMKEHAKVVOQWRIRTESEVTADALATIDGLIGDSERLT- 59
 Db 1 MLLVDFVGNTHVTGLGFDG----DKLVNWRIKTDREKTSDEYGLILSNLFDYDNVNSD 56
 Qy 60 --GTAALSTVPSVLHEVRIMLDQWPSVPHVLTPEGVRTGIPLLVDNPKKEVGADRIVNC 117
 Db 57 IDVDIISVVPVNMVHSLNFECIKYCKQP-LIVPGIKTGINKYDNPQVGDRIYNAV 115
 Qy 118 AAYDRFKAAIIVDFGSSICVDVVSAGKEFLGGAIAFGVQVSSDAAAARSALRRVELAR 177
 Db 116 AGIEKGAPSLVDFGTATTFCALISERGEYLGGTIAPGIIKISSEALFQSASKLPRVELAK 175
 Qy 178 PRSVGNKNTVECHQAGAVFGAGLVGDLVGRIRREDVSGFSDHDAIVATGHTAPLLE 237
 Db 176 PGMTICKSTVSAMQSGIIVYGVGLVDKIIISIMKELN---CDDVKVIATGGLAKLIASE 231
 Qy 238 LHTVDHYDQHLTLOGLRLVFERNLE 262
 Db 232 TKSIDYVDGFLTLEGLRIIYEKNOE 256

RESULT 5
 AAU91154
 ID AAU91154 standard; Protein; 255 AA.
 XX AC AAU91154;
 XX
 XX 05-JUN-2002 (first entry)
 XX Geobacter sulfurreducens pantothenate kinase Coax.
 DE Geobacter sulfurreducens pantothenate kinase Coax.
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX Geobacter sulfurreducens.
 XX
 XX WO200216601-A2.
 XX
 XX 28-FEB-2002.
 XX
 XX 24-AUG-2001; 2001WO-US26531.
 PF
 XX 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 PR

XX Claim 10; Page 70-71; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX SQ Sequence 272 AA;

Query Match 100.0%; Score 1367; DB 23; Length 272;

Best Local Similarity 100.0%; Pred. No. 1.2e-129;

Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVQVQWRIETSEVTADELALTDGLIGEDSERLTG 60

1 MLLAIDVRNTHTVVGLSGMKEHAKVQVQWRIETSEVTADELALTDGLIGEDSERLTG 60

QY 61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

Db 61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

QY 121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

Db 121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

QY 181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

Db 181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

QY 241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

Db 241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

RESULT 2

AAG81225

ID AAG81225 standard; Protein; 272 AA.

XX AAG81225;

XX 04-SEP-2001 (first entry)

DT Mycobacterium tuberculosis potential drug target protein SEQ ID 276.

XX Drug target; growth; organism viability; characterisation.

XX Mycobacterium tuberculosis.

XX WO200135317-A1.

XX 17-MAY-2001.

XX 13-NOV-2000; 2000WO-US31152.

XX 12-NOV-1999; 99US-0165086.

XX 12-NOV-1999; 99US-0165124.

XX 01-FEB-2000; 2000US-0179531.

XX (RBGC) UNIV CALIFORNIA.

XX Eisenberg D, Rotstein SH, Marcotte EM;

XX WPI; 2001-329193/34.

XX N-PSDB; AAH52076.

XX Identifying nucleotide or polypeptide sequence for use as drug target,

PT Involves providing algorithm that analyzes a functional relationship

PT between nucleotide or polypeptide sequences, and comparing the

PT sequences

XX Disclosure; Page 188; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or

CC polypeptide sequence that may be a drug target, or essential for growth

CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium

CC tuberculosis proteins which are potential drug targets. The DNA and

CC protein sequences are used to illustrate the method of the invention. The

CC method involves providing an unknown nucleotide or polypeptide sequences,

CC and comparing it to a number of sequences along with at least one

CC algorithm capable of analysing a functional relationship between

CC nucleotide and polypeptide sequences. The method is useful for

CC characterising the function of nucleic acids and polypeptides that may be

CC useful as a target for a drug or essential for the growth or viability of

CC an organism.

XX SQ Sequence 272 AA;

Query Match 99.7%; Score 1363; DB 22; Length 272;

Best Local Similarity 99.6%; Pred. No. 3.1e-129;

Matches 271; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLAIDVRNTHTVVGLSGMKEHAKVQVQWRIETSEVTADELALTDGLIGEDSERLTG 60

1 MLLAIDVRNTHTVVGLSGMKEHAKVQVQWRIETSEVTADELALTDGLIGEDSERLTG 60

Db 1 MLLAIDVRNTHTVVGLSGMKEHAKVQVQWRIETSEVTADELALTDGLIGEDSERLTG 60

61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

Db 61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

61 TAALSTVPSVLHEVRIMLDQWPSVPHVLTIEPGVRTGIPLLVDPNPKVGVGADRVNCLAA 120

QY 121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

Db 121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

121 DRFRKAAIVVDFGSSICVDVVSAGFELGCAIAPGVQVSSDAAAARSALRRVELARPRS 180

QY 181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

Db 181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

181 VVGKNTVECMQAGAVFGAGLVGRLVGRIRREDVSGFSVDHDAIVATGHTAPLPPH 240

QY 241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

Db 241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

241 VDHYDQHLTLQGLRLVFERNLEVQGRGLTKAR 272

RESULT 3

AAG91151

ID AAG91151 standard; Protein; 265 AA.

XX AAG91151;

XX 05-JUN-2002 (first entry)

DT Streptomyces coelicolor pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Streptomyces coelicolor.

XX WO2002:6601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 27.9257 seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-5
Perfect score: 1367
Sequence: 1 MLLADVNTHTVTVGLSGM.....LRLVFNLEVRQGRLLKTAR 272

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1367	100.0	272	AAU91152	Mycobacterium tube
2	1363	99.7	272	AA181225	Mycobacterium tube
3	698.5	51.1	265	AAU91151	Streptomyces coeli
4	513	37.5	256	AAU91175	Pantothenate kinas
5	509.5	37.3	255	AAU91154	Geobacter sulfurre
6	494.5	36.2	258	AAU01243	B. subtilis novel
7	494.5	36.2	258	AAU91149	Bacillus subtilis
8	487.5	35.7	254	AAU91171	Pantothenate kinas
9	479.5	35.1	258	AAU91172	Pantothenate kinas
10	471.5	34.5	262	AAU91170	Pantothenate kinas

11	456.5	33.4	259	23	ABB47661	Listeria monocytog
12	431.5	31.6	258	23	AAU91153	Rhodobacter capsul
13	412	30.1	233	23	AAU91163	Pantothenate kinas
14	404	29.6	262	23	AAU91155	Deinococcus radiop
15	389	28.5	250	23	AAU91150	Clostridium acetob
16	382.5	28.0	260	23	AAU91173	Pantothenate kinas
17	371	27.1	219	23	AAU91176	Pantothenate kinas
18	286	20.9	246	23	AAU91156	Thermotoga maritim
19	274.5	20.1	257	23	AAU91174	Pantothenate kinas
20	254	18.6	273	23	AAU91157	Treponema pallidum
21	209.5	15.3	212	23	AAU91177	Pantothenate kinas
22	160.5	11.7	249	23	AAU91178	Pantothenate kinas
23	159	11.6	241	23	AAU91179	Pantothenate kinas
24	153	11.2	244	23	AAU91168	Pantothenate kinas
25	152	11.1	455	21	AAU74910	Neisseria meningit
26	152	11.1	592	20	AAU38616	Neisseria meningit
27	152	11.1	592	21	AAU74913	Neisseria meningit
28	152	11.1	592	23	AAU91166	Pantothenate kinas
29	150.5	11.0	249	23	AAU91182	Pantothenate kinas
30	149	10.9	248	23	AAU91164	Pantothenate kinas
31	148	10.8	389	21	AAU74909	Neisseria meningit
32	147.5	10.8	592	20	AAU38615	Neisseria meningit
33	147.5	10.8	592	21	AAU74912	Neisseria meningit
34	147.5	10.8	592	23	AAU91169	Pantothenate kinas
35	146.5	10.7	455	20	AAU38617	Neisseria gonorrh
36	146.5	10.7	455	21	AAU74908	Neisseria gonorrh
37	146.5	10.7	460	23	AAU91167	Pantothenate kinas
38	146.5	10.7	592	20	AAU38618	Neisseria gonorrh
39	146.5	10.7	592	21	AAU74911	Neisseria gonorrh
40	138	10.1	262	23	AAU91158	Borrelia burgdorfe
41	119	8.7	338	20	AAU29123	Amino acid sequenc
42	116.5	8.5	189	20	AAU38614	Neisseria meningit
43	114	8.3	229	23	AAU91159	Aquifex aeolicus p
44	112	8.2	242	23	AAU91180	Pantothenate kinas
45	106	7.8	257	23	AAU91160	Synechocystis pant

ALIGNMENTS

RESULT 1
AAU91152
ID AAU91152 standard; Protein; 272 AA.

AC AAU91152;

DT 05-JUN-2002 (first entry)

DE Mycobacterium tuberculosis pantothenate kinase Coax.

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.

OS Mycobacterium tuberculosis.

PN WO200216601-A2.

PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26531.

PR 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

PR (OMNI-) OMNIGENE BIOPRODUCTS INC.

PI Yocum RR, Patterson TA;

DR WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises
contacting composition comprising pantothenate kinase (Coax) protein
with test compound and identifying inhibitor of the Coax protein -

R;Kunst, F.; Ogasawara, N.; Meszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertel
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallen
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.P.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69740
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
C;Cross-references: GB:299104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:e1182003;
Experimental source: strain 168
C;Genetics:
A;Gene: yacB
A;Start Codon: TTG
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 30.1%; Score 412; DB 2; Length 233;
Best Local Similarity 39.4%; Pred. No. 2.6e-25;
Matches 84; Conservative 45; Mismatches 76; Indels 8; Gaps 3;
QY 1 MLLAIDVRNTHVTVGL--LSCMKHAKVQVQWRIRTESEVTADELALTDIGLIGDS---ER 57
Db 1 MLLAIDVRNTHVTVGL--LSCMKHAKVQVQWRIRTESEVTADELALTDIGLIGDS---ER 57
QY 58 LTGTAALSTVPSVLHVRIMLDQYWPSPVHLIEPCVGTGIPLLVDPNPKVGVGADRVNCL 117
Db 57 IDGIISSVPPVPMFALERTCTKFIHIEPOI-VGPMKTKGLNFKYDNPKVGVGADRVNCL 115
QY 118 AAYDRFRKAAIVVDFGSSICDVVSAAKGFGLGAIAPGVQVSSDAAAARSALRRVELAR 177
Db 116 AATHLGNPLVVDGTAITYCYIDENKQYMGGAIPGITISTEALYSRAAKLPRIEIR 175
QY 178 PRSVGKNTVECMQAGAVFGAGLVGLVRIR 210
Db 176 PDNIIGKNTVSAMQSGILFGYVGVQVGVKRMK 208
RESULT 9
5516
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)
C;Species: *Deinococcus radiodurans*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C;Accession: E75516
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <WHI>
C;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g645814
Experimental source: strain R1
C;Genetics:
A;Gene: DR0461
A;Map position: 1
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 29.6%; Score 404; DB 2; Length 262;
Best Local Similarity 38.3%; Pred. No. 1.3e-24;

Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;
QY 2 LLAIDVRNTHVTVGL--LSCMKHAKVQVQWRIRTESEVTADELALTDIGLIGDSERLT 59
Db 6 LLAVIDGNITTVGLADASSALHTHT-----WRITRNREMLPDDLAQLHGLTLAGAPIP 60
QY 60 GTAALSTVPSVLHVRIMLDQYWPSPVHLIEPCVGTGIPLLVDPNPKVGVGADRVNCL 109
Db 61 RAAVLSSVAPPVGENYALAKRHFMDAFASAEANLPDVTVE-----LDTPGSVG 110
QY 110 ADRIIVNCLAA--YDRFKAAIVVDFGSSICDVVSAAKGFGLGAIAPGVQVSSDAAAARS 167
Db 111 ADRLCMLFGAEKYLGLDVAVVDFGTNFNDVVGRRPLGILATGAQVSADALFARA 170
QY 168 AALRRVELARPSVGVKNTVECMQAGAVFGAGLVGLVRIRTESEVTADELALTDIGLIGDS 227
Db 171 AKLPRLITLQAPETAIGKNTVHALQSLGVGAEMVDGLLRIRAEPLG-----EAVAVAT 225
QY 228 GHTAPLLLPDLHVTVDYDQHLILQGLRLVFERNLEVO 264
Db 226 GGFSTVQVIGQIDYIDETLTGLGLVELWASRSEVR 262
RESULT 10
B87489
transcription activator, probable Baf family [imported] - *Caulobacter crescentus*
C;Species: *Caulobacter crescentus*
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
C;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 28.0%; Score 382.5; DB 2; Length 261;
Best Local Similarity 34.0%; Pred. No. 6.3e-23;
Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;
QY 1 MLLAIDVRNTHVTVGLLSCMKHAKVQVQWRIRTESEVTADELALTDIGLI---GEDSER 57
Db 2 MLLAIEQGNNTMFALHDG-----ASWVAQWRSATSTRTADEYVWVWLSQLLSMOGLGFRA 57
QY 58 LTGTAALSTVPSVLHVRIMLDQYWPSPVHLIEPCVGTGIPLLVDPNPKVGVGADRVNCL 117
Db 58 IDAVIISVVPOSIFNRLNLSRYF-NVEPLVIGENAKLIDIVRIKPEAGADRVNCL 116
QY 118 AAYDRFRKAAIVVDFGSSICDVVSAAKGFGLGAIAPGVQVSSDAAAARSALRRVELAR 177
Db 117 GAAMVPGPLVWIDSTATTDFDAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIOR 176
QY 178 P-RSVVGNKNTVECMQAGAVFGAGLVGLVRIRTESEVTADELALTDIGLIGDSERLT 235
Db 177 PAGNRIVGTDVTSAMQSGVFWGYSILIEGLVARIKAEK-----EPMTVIATGGVASLFE 231
QY 236 PELHTVDYDQHLITLQGLRLVFERN 260
Db 232 GATDSIDHFDSDLTIRGLLEIYARN 256
RESULT 11
D72320
conserved hypothetical protein - *Thermotoga maritima* (strain MSB8)
C;Species: *Thermotoga maritima*

DB	113	NAVAASEEGTPIVVDFCTATFCYIDESGVYQGGAITAPGIMISTEALYNRAAKLPRVD	172
QY	175	LARPRSVYGKNTVECMQAGAVGFAGLVGLVGRIRIEDVSGFSVDHVAIVATGHTAPLL	234

175 LARPSVWGKNTVECHQAGAVFGAGLVGLGRIREDVSGFSDHDAIVATGHTAPLL 234

A:Residues: 1-233 <OGA>
A:Cross-references: EMBL:D26185; NID:g467326; PID:BAA05305.1; PID:d1005847; PID:g467326;
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1999.

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H86937
R;Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
C;Accession: H86937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:gl3092576; PIDN:CAC29740.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0232
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 90.9%; Score 1242; DB 2; Length 274;
Best Local Similarity 90.9%; Pred. No. 8.4e-91;
Matches 249; Conservative 9; Mismatches 14; Indels 2; Gaps 1;
QY 1 MLIAIDVRNTHVTGVLSSGMEKHAQVQWRIETSEVTADELALITDGLIGDSERLIG 60
DB 1 MLIAIDVRNTHVTGVLSSGMEKHAQVQWRIETSEVTADELALITDGLIGDSERLAG 60
QY 61 TAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVGGADRVNCLAA 120
DB 61 AAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVGGADRVNCLAA 120
QY 121 DFRKAAIYVDFSSICVDVWSAKGEFLGGAIAAGVQVSSDAAAARLRVVELAPRS 180
DB 121 HKFGQAAIYVDFSSICVDVWSAKGEFLGGAIAAGVQVSSDAAAARLRVVELAPRS 180
QY 181 VVGKNTVECMQAGAVFAGLVGRIREDVSGFSDV--HDAIVATGHTAPLLPEL 238
DB 181 VVGKNTVECMQAGAVFAGLVGRIREDVSGFSDV--HDAIVATGHTAPLLPEL 240
QY 239 HTVDHYDQHLTLQGLRLVFERNL 272
DB 241 HTVDHYDQHLTLQGLRLVFERNL 274
RESULT 3
hypothetical protein SCE94.31c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
Accession: T36391
Submitted to the EMBL Data Library, April 1999
A;Reference number: 221573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOE94.31c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 51.1%; Score 698.5; DB 2; Length 265;
Best Local Similarity 51.5%; Pred. No. 7e-48;
Matches 138; Conservative 53; Mismatches 64; Indels 13; Gaps 3;
QY 1 MLIAIDVRNTHVTGVLSSGMEKHAQVQWRIETSEVTADELALITDGLIG-----E 53
DB 1 MLITIDVGNTHVTGLGFDG---EDIVEHWRISTDSRRTADELAVLLQGLMGHMLLIGDE 56
QY 54 DSERLNCTAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVGGADRI 113
DB 57 LGGIDGIACTAVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVGGADRI 116

QY 114 VNCLAAIDYRFRKAAIYVDFSSICVDVWSAKGEFLGGAIAAGVQVSSDAAAARLRV 173
DB 117 INAAVAVELYGGPAIYVDFGTTTDFAVSARGEYIGVIAPIGIEISVEALGVKGAQLRKI 176
QY 174 ELARPSRVGKNTVECMQAGAVFAGLVGRIREDVSGFSDVHDVAIVATGHTAPL 233
DB 177 EVARPSRVGKNTVECMQAGAVFAGLVGRIREDVSGFSDVHDVAIVATGHTAPL 234
QY 234 LLPELHTVDHYDQHLTLQGLRLVFERNL 261
DB 235 VLGESSVIDHEPWLTLMLGLRLVIERNV 262

RESULT 4

F83660
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83660
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:gl0172612; PIDN:BA03805.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BHC086
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 35.7%; Score 487.5; DB 2; Length 254;
Best Local Similarity 37.9%; Pred. No. 3.1e-31;
Matches 100; Conservative 64; Mismatches 87; Indels 13; Gaps 4;
QY 1 MLIAIDVRNTHVTGVLSSGMEKHAQVQWRIETSEVTADELALITDGLI---GEDSER 57
DB 1 MILVIDGNTNTVLGY---QDETIVHWRLATSRQKTEDEYAMTVRSIFDHAGLQFD 56
QY 58 LGCTAALSTVPSVLHVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVGGADRVNCL 117
DB 57 IDGIVISSVPPMFMFSLQCKKYFHTP--MIIGPGIKTGLNPKYDNPKVGGADRVN 115
QY 118 AAYDRFRKAAIYVDFSSICVDVWSAKGEFLGGAIAAGVQVSSDAAAARLRVVELAR 177
DB 116 AAELGYPAIYVDFGTTTDFAVSARGEYIGVIAPIGIEISVEALGVKGAQLRKI 175
QY 178 PRSVGKNTVECMQAGAVFAGLVGRIREDVSGFSDVHDVAIVATGHTAPLLPE 237
DB 176 PRQVCTNTIDSMQSGIFGYGVQVSDGVVKRMA----QAEEPKVIATGGKLIGTE 230
QY 238 LHTVDHYDQHLTLQGLRLVFERNL 261
DB 231 SETIDVIDSFLTKGLQLIYKKNV 254

RESULT 5

E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostr
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A95900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.92097 seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-5
Perfect score: 1367
Sequence: 1 MLLAIDVRNTHVVGLLSGM.....LRLVFNLEVQGRGLKTKAR 272

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1367	100.0	272	2 A70955	hypothetical prote
2	1242	90.9	274	2 H86937	conserved hypotet
3	698.5	51.1	265	2 T36391	hypothetical prote
4	487.5	35.7	254	2 F83660	hypothetical prote
5	467.5	34.2	273	2 E97293	probable transcrip
6	456.5	33.4	259	2 AF1102	conserved hypotet
7	448.5	32.8	259	2 AF1464	conserved hypotet
8	412	30.1	233	2 S66100	conserved hypotet
9	404	29.6	262	2 E75516	conserved hypotet
10	382.5	28.0	261	2 B87489	transcription acti
11	286	20.9	246	2 D72320	conserved hypotet
12	254	18.6	273	2 D71326	conserved hypotet
13	152	11.1	592	2 H82031	probable biotin-la
14	149	10.9	248	2 H83111	hypothetical prote
15	147.5	10.8	592	2 B81009	BirA protein/Bvg a
16	138	10.1	262	2 F70165	conserved hypotet
17	114	8.3	229	2 E70465	hypothetical prote
18	112	8.2	242	2 A82637	conserved hypotet
19	108.5	7.9	543	2 S73595	heat shock protein
20	106	7.8	257	2 S73559	hypothetical prote
21	101	7.4	732	2 A90863	cell division cont
22	100.5	7.4	264	2 C70603	hypothetical prote
23	100	7.3	255	2 AG1744	conserved hypotet
24	100	7.3	276	2 A12292	hypothetical prote
25	97.5	7.1	223	2 G71887	hypothetical prote
26	97.5	7.1	1246	2 J00406	hypothetical prote
27	97.5	7.1	4735	2 T17463	rifamycin polyketi
28	97	7.1	473	2 H97426	flagellum-specific
29	97	7.1	473	2 AH2644	flagellum-specific

30	96.5	7.1	597	2 C69283	hypothetical prote
31	96	7.0	730	2 E75387	NADH dehydrogenase
32	94.5	6.9	255	2 A11374	conserved hypotet
33	93.5	6.8	746	2 E83250	still frameshift 3
34	93	6.8	405	2 T47595	RING finger protei
35	92.5	6.8	267	2 I40327	baf protein - Bord
36	92.5	6.8	465	2 A84176	mercury(II) reduct
37	92.5	6.8	543	2 D64243	heat shock protein
38	92	6.7	433	2 S74407	homoserine dehydro
39	92	6.7	462	2 T27323	hypothetical prote
40	91.5	6.7	209	2 H81382	hypothetical prote
41	91	6.7	391	2 E87449	conserved hypotet
42	91	6.7	449	2 E87639	hypothetical prote
43	91	6.7	471	2 AB3422	glutamyl-tRNA(gln)
44	89.5	6.5	223	2 F64627	hypothetical prote
45	89	6.5	337	2 S60577	porphobilinogen sy

ALIGNMENTS

RESULT 1

A70955
hypothetical protein RV3600c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70955
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70955
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-272 <CDS>
A:Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g21133
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3600c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	100.0%	Score 1367;	DB 2;	Length 272;
Best Local Similarity	100.0%	Pred. No. 1.1e-100;	Mismatches 0;	Gaps 0;
Matches 272;	Conservative 0;			
Qy	1	MLLAIDVRNTHVVGLLSGMKEHAKVVOQWRIETSEVTADELALTIDGLIGEDSERLTG	60	
Db	1	MLLAIDVRNTHVVGLLSGMKEHAKVVOQWRIETSEVTADELALTIDGLIGEDSERLTG	60	
Qy	61	TAALSTVPSVLHVRIMLDQYWPSPHYLIEPGVRTGTPLLVDNPKVGADRVNCLAA	120	
Db	61	TAALSTVPSVLHVRIMLDQYWPSPHYLIEPGVRTGTPLLVDNPKVGADRVNCLAA	120	
Qy	121	DRFKAAIVDPFGSSICVDVVSAGFELGGAIAPGVQVSSDAAAARSALRRVELARPRS	180	
Db	121	DRFKAAIVDPFGSSICVDVVSAGFELGGAIAPGVQVSSDAAAARSALRRVELARPRS	180	
Qy	181	VVGNTVBCMAGAVFGFAGLVGVRIRDEVSGFSDHVDVAIVATGHTAPLLPELHT	240	
Db	181	VVGNTVBCMAGAVFGFAGLVGVRIRDEVSGFSDHVDVAIVATGHTAPLLPELHT	240	
Qy	241	VDHYDQHLTLQGLRLVFNLEVQGRGLKTKAR	272	
Db	241	VDHYDQHLTLQGLRLVFNLEVQGRGLKTKAR	272	

RESULT 2

H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae

THIS PAGE BLANK (USPTO)

```
QY 61 TAAL---STVPSVLHEVRIMLDQWPSVPHVLIEPGVRTGIPLLVD---NPKEVGADRIY 114
Db 1 : | | : : : : | : : : : | : | | | |
58 VDAILISSVVPNIITFTQFARKYF-KVEATIVD--LEKKLPFTFAKGINVTGFCADRII 114
QY 115 NCLAAAYDRF-RKAAIWDVFGSSICVDVVSAGKEFLGGAIAPGVQVSSDAAARSALRRV 173
Db 1 : | | : : : : | : | | | | : | : | | |
115 DITEAMQYIPKLNLIFFDGTATVDVLK-RGVYIGGGILPGIDMSINALYGNATKLRPV 173
QY 174 ELARPSVVGKNTVECMAGAVFGFAGLVGDLVGRVREDVSGFSVDHDAIVATGHTAPL 233
Db 1 : | | : : : : | : | | | | : | : | | |
174 KFTPTSSVIGDITMQAIGQAFVAGQIKHIIKINEE-----LNEEIFVLATGGLGKI 228
QY 234 LLPELHTVDYDHOHLTLOGLRVFERN 260
Db 1 : | | : : : : | : | | | | : | : | | |
229 LSAEIDEIDEYDANLSLGLVLYKLN 255

RESULT 14
Q9WZY5 PRELIMINARY: PRT: 246 AA.
Q9WZY5;
01-NOV-1999 (TremBLrel. 12, Created)
01-NOV-1999 (TremBLrel. 12, Last sequence update)
01-JUN-2002 (TremBLrel. 21, Last annotation update)
Hypothetical protein TW0883.
GN TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS88 / DSM 3109;
RA MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
"Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -.
DR TIGR; TM0883; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 20.9%; Score 286; DB 16; Length 246;
Best Local Similarity 31.4%; Pred. No. 4.2e-15;
Matches 80; Conservative 49; Mismatches 108; Indels 18; Gaps 7;
QY 1 MLLAIDVRNTHTVVGLLSGMKEHAKVVOQWRIRTESEVTADALITDGLIGEDSERLTG 60
Db 1 : | | : : : : | : | | | | : | : | | |
1 MYLLVDGNTSHVSFI----TEDGKTFRRRLSTGVTFQTEDELSHLPGLGDAMREIKG 56
QY 61 TAALSTVPSVLHEVRIMLDQWPSVPHVLIEPGVRTG--IPLVDNPKVEGADRIYNCLAA 119
Db 1 : | | : : : : | : | | | | : | : | | |
57 IGVASVWPNTQNTIERFSQKVFHISP----IWKAKNGCVKWNKPNPSEVGADRIYANVAF 113
QY 120 YDRFRKAAIIVDFSSICVDVVSAGKEFLGGAIAPGVQVSSDAAARSALRRVELARPR 179
Db 1 : | | : : : : | : | | | | : | : | | |
114 VKEGKNGIIDMGATTTVDLV-VNGSYEGAILPGFMVHSLFRGTAKLPLVEKPAD 172
QY 180 SVWGNKTVCEMQAGAVFGFAGLVGDLVGRVREDVSGFSVDHDAIVATGHTAPL 238
Db 1 : | | : : : : | : | | | | : | : | | |
173 FVVGKDTENIRLGVNVSVALEGIIGRIKE-----VYGLDPVLVTGGQSKIVKDIK 226
QY 239 HTVDHYDHOHLTLOGL 253
Db 1 : | | : : : : | : | | | | : | : | | |
227 HEI--FDIEDLTIKGV 239
```

```
RESULT 15
O83446 PRELIMINARY: PRT: 273 AA.
ID O83446
AC O83446;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein TP0431.
GN TP0431.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RA MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
Venter J.C.;
"Complete genome sequence of Treponema pallidum, the syphilis
spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC65417.1; -.
DR TIGR; TP0431; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 18.6%; Score 254; DB 16; Length 273;
Best Local Similarity 31.3%; Pred. No. 1.7e-12;
Matches 84; Conservative 42; Mismatches 118; Indels 24; Gaps 7;
QY 2 LLAIDVRNTHTVVGLLSGMKEHAKVVOQWRIRTESEVTADALITDGL-----IGEDSE 56
Db 1 : | | : : : : | : | | | | : | : | | |
1 MLLIDVGNSHVVFQIGENGGRVVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60
QY 57 RLGTGTAALSTVPSVLHEVRIMLDQWPSVPHVLIEPGVRTGIPLLVDNP--KEVGADRIY 114
Db 1 : | | : : : : | : | | | | : | : | | |
61 R---DAFTSSVVPVLTKTITADAVAQISGVQVPGVFWAYEHLFVRPEPVRABIGTDLVA 117
QY 115 NCLAAAYDRFRKAAIIVDFSSICVDVVSAGKEFLGGAIAPGVQVSSDAAARSALRRVE 174
Db 1 : | | : : : : | : | | | | : | : | | |
118 NAAAYVHFRSACVVDGCGTALTFTAVDGTGLIOGVATAPGLRTAVQSILHTGTAQLPLVP 177
QY 175 LARPSVVGKNTVECMAGAVFGFAGLVGDLVGRVREDVSGFSVDHDAIVATGHTAPL 234
Db 1 : | | : : : : | : | | | | : | : | | |
178 LALPDSVLGKDTTHAVQAGVVRGTILFVIRAMIAQCQKELG-----CRCAAVITGSLRFL 232
QY 235 LPELHTVDH--YDQHLTLOGL----RLV 256
Db 1 : | | : : : : | : | | | | : | : | | |
233 SSE---VDFPPIDAQLTSLSLAHIAIRLV 257

Search completed: June 24, 2003, 21:59:18
Job time : 24.1844 secs
```


Db 187 PAYAICKNTISSQGIYGVYIVRMKEELQE-EGEKEPLVATGGLAKLISEE 245
 * QY 238 LHTVDHYDOHLTQGLRLVFN 260
 Db 246 AKNDVINPFLTLEGLRIYKFN 268

* RESULT 9

Q8YAC5 PRELIMINARY; PRT; 259 AA.
 AC Q8YAC5;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical protein lmo0221.
 GN LMO0221.
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-E / SEROVAR 1/2a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlant J., Cossart P.,
 RT "Comparative genomics of Listeria species."
 RL Science 294:849-852(2001).
 DR EMBL; AL591974; CAD00748.1; -.
 DR MEROPS; M1.009; -.
 DR ListList; LMO00221; -.
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMs; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 33.4%; Score 456.5; DB 16; Length 259;
 Best Local Similarity 36.5%; Pred. No. 9.9e-29;
 Matches 97; Conservative 60; Mismatches 90; Indels 19; Gaps 5;
 QY 1 MLLAIDVRNTHVTVGLSGMKEHAKVQVQWRIRTESEVTADELALTI-----DGLIGED 54
 Db 1 MLIVDVGNTCTGVY----EKQLLRHWRMTDRHRTSDELGTVLNFYSYANLTPSD 56
 QY 55 SERLTGTAALSTVPSVLHEVRIMLDQYWPSPVHLIEPCVRGIPGLLVNDPKVEGADRV 114
 Db 57 ---IQGIISVVPPIMHAMETMCVRYF-NIRPLIVGPGIKTGLNPKVDNPREIGSDRV 112
 QY 115 NCLAAIDRFRKAAIIVDFGSSICVDVSAKGEFLGGAIPGVQVSSDAAAARSALRRVE 174
 Db 113 NAVAASEYGPVIVDFGTATTCYIDEGVYQGGATAPGIMISTEALYNRAKLPRVD 172
 QY 175 LARPSVVGKNTVECMQAGVFGAGLVGRLREDVSGFSVDHDAIVATGHTAPLL 234
 Db 173 IAESSQIGTKSVASMQAGIFYGFTGCEGIIAEMKQSNTPV-----VATGGLARMI 227
 QY 235 LPELHTVDHYDOHLTQGLRLVFN 260
 Db 228 TEKSADVILDPFLTKGLELLYRN 253
 QY 175 LARPSVVGKNTVECMQAGVFGAGLVGRLREDVSGFSVDHDAIVATGHTAPLL 234
 Db 173 IAESSQIGTKSVASMQAGIFYGFTGCEGIIAEMKQSNTPV-----VATGGLARMI 227
 QY 235 LPELHTVDHYDOHLTQGLRLVFN 260
 Db 228 TEKSADVILDPFLTKGLELLYRN 253

* RESULT 10

Q8YAC5 PRELIMINARY; PRT; 262 AA.
 AC Q8YAC5;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical protein DR0461.
 GN DR0461.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;

Q92F54 PRELIMINARY; PRT; 259 AA.
 AC Q92F54;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical protein lln0253.
 GN Lln0253.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / SEROVAR 6A;
 RX PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlant J., Cossart P.,
 RT "Comparative genomics of Listeria species."
 RL Science 294:849-852(2001).
 DR EMBL; AL596164; CAC95486.1; -.
 DR ListList; Lln00253; -.
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMs; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;
 Query Match 32.8%; Score 448.5; DB 16; Length 259;
 Best Local Similarity 36.1%; Pred. No. 4.3e-28;
 Matches 96; Conservative 61; Mismatches 90; Indels 19; Gaps 5;
 QY 1 MLLAIDVRNTHVTVGLSGMKEHAKVQVQWRIRTESEVTADELALTI-----DGLIGED 54
 Db 1 MLIVDVGNTCTGVY----KEQKLLRHRMTDRHRTSDELGTVLNFYSYANLTPSD 56
 QY 55 SERLTGTAALSTVPSVLHEVRIMLDQYWPSPVHLIEPCVRGIPGLLVNDPKVEGADRV 114
 Db 57 ---IQGIISVVPPIMHAMETMCVRYF-NIRPLIVGPGIKTGLNPKVDNPREIGSDRV 112
 QY 115 NCLAAIDRFRKAAIIVDFGSSICVDVSAKGEFLGGAIPGVQVSSDAAAARSALRRVE 174
 Db 113 NAVAASEYGPVIVDFGTATTCYIDEGVYQGGATAPGIMISTEALYNRAKLPRVD 172
 QY 175 LARPSVVGKNTVECMQAGVFGAGLVGRLREDVSGFSVDHDAIVATGHTAPLL 234
 Db 173 IAESSQIGTKSVASMQAGIFYGFTGCEGIIAEMKQSNTPV-----VATGGLARMI 227
 QY 235 LPELHTVDHYDOHLTQGLRLVFN 260
 Db 228 TEKSADVILDPFLTKGLELLYRN 253
 RESULT 11
 Q9RX54 PRELIMINARY; PRT; 262 AA.
 AC Q9RX54;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical protein DR0461.
 GN DR0461.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;

Db	127	AAHEIYKRSLIIIDFGTATTC	CAVRENGYLGGAICPGIKVSSEALFEAKALPRVELIK	186
QY	178	PRSVGKNTVECMQAGAVFGFAGLVDGLVGRIR	EDVSGFSGVDHDAIVAIGTHTAPLLPPE	237

```
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match
Best Local Similarity 51.1%; Score 698.5; DB 16; Length 265;
Matches 138; Conservative 53; Mismatches 64; Indels 13; Gaps 3;

QY 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLIG-----E 53
DB 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLIG-----E 53
QY 54 DSELTGTAALSTVPSVLHVRIMLDQYWPSPVPHVLEPGVTRTGPILLVDNPKVEGADRI 113
DB 57 LGDGDGTAICATPVSILHRELVTRRYIGDVPVAVLPEGVKGTGPILLTDHPKEVGADRI 116
QY 114 VNCLAAVDRFRKAAIIVDFGSSICVDVVSARKGEFLGGAIAIPGVQVSSDAAAARSAAALRRV 173
DB 117 INAAVAVELYGGPAIVDFGATTFDAVSARGEYIGGVIAPIEISVEALGVKGALRKI 176
QY 174 ELAPRSVGNKTVCEMQAGAVFGAGLVGDLGVRIRDSVGFSDHVDVAIVATGHTAPL 233
DB 177 EVARPSVIGNTVEAMQSGIVYGFAGQVGVNRMARELA--DDPDVTVIATGGLAPM 234
QY 234 LLPELTHVDYDQHLTLOGLRLVFERNL 261
DB 235 VLGESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 4
Q8XHL5 PRELIMINARY; PRT; 259 AA.
AC Q8XHL5
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DE Hypothetical protein CPE2468.
GN CPE2468
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
"Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB82174.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

Query Match
Best Local Similarity 38.1%; Score 520.5; DB 16; Length 259;
Matches 102; Conservative 73; Mismatches 81; Indels 9; Gaps 4;

QY 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLICE---DSE 57
DB 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLICE---DSE 57
QY 58 LTGTAALSTVPSVLHVRIMLDQYWPSPVPHVLEPGVTRTGPILLVDNPKVEGADRI 117
DB 57 VEGIIISVVPNIMHLENMVRKCFCKEP-IVVGGIKTGINKYDNPKEVGADRI 115

Pfam: PF03309; Bvg_acc_factor; 1.
TIGRFAMs: TIGR00671; baf; 1.
Hypothetical protein.
SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match
Best Local Similarity 51.5%; Pred. No. 4.3e-48;
Matches 138; Conservative 53; Mismatches 64; Indels 13; Gaps 3;

QY 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLIG-----E 53
DB 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLIG-----E 53
QY 54 DSELTGTAALSTVPSVLHVRIMLDQYWPSPVPHVLEPGVTRTGPILLVDNPKVEGADRI 113
DB 57 LGDGDGTAICATPVSILHRELVTRRYIGDVPVAVLPEGVKGTGPILLTDHPKEVGADRI 116
QY 114 VNCLAAVDRFRKAAIIVDFGSSICVDVVSARKGEFLGGAIAIPGVQVSSDAAAARSAAALRRV 173
DB 117 INAAVAVELYGGPAIVDFGATTFDAVSARGEYIGGVIAPIEISVEALGVKGALRKI 176
QY 174 ELAPRSVGNKTVCEMQAGAVFGAGLVGDLGVRIRDSVGFSDHVDVAIVATGHTAPL 233
DB 177 EVARPSVIGNTVEAMQSGIVYGFAGQVGVNRMARELA--DDPDVTVIATGGLAPM 234
QY 234 LLPELTHVDYDQHLTLOGLRLVFERNL 261
DB 235 VLGESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 5
Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
DE TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
Tan H., Chen R., Wang J., Yu J., Yang H.;
"A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013180; AAM25520.1; -
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620EC8C8A6ED CRC64;

Query Match
Best Local Similarity 36.8%; Score 503.5; DB 16; Length 255;
Matches 105; Conservative 64; Mismatches 83; Indels 13; Gaps 4;

QY 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLIGDSERLT- 59
DB 1 MLLAIDVRNTHVTVVGLSGMKEHAKVVOQWRIRTESEVTADELATIDGLIGDSERLT- 59
QY 60 --GTAALSTVPSVLHVRIMLDQYWPSPVPHVLEPGVTRTGPILLVDNPKVEGADRI 117
DB 57 IDDVIISSVVPPLMTLQVMSLKVFRTKP-IVVGGIKTGINKYDNPKEVGADRI 115
QY 118 AAYDRFRKAAIIVDFGSSICVDVVSARKGEFLGGAIAIPGVQVSSDAAAARSAAALRRV 177
DB 116 AAYELGPGVIVDFGATTFCAISEKGEGLGIIAPGLMISADALFORTAKLPKIDLT 175
QY 178 PRSVGKNTVCEMQAGAVFGAGLVGDLGVRIRDSVGFSDHVDVAIVATGHTAPL 237
DB 176 PPTVINRNTVASMOSGIYGVHGMVDYIVTRMKGEFA-----PSAYVYVATGTFANMIAE 230
QY 238 LHTVDHYDQHLTLOGLRLVFERNLE 262
DB 231 SKTIDTVNEMTLLEGRLIYERNKE 255

RESULT 6
Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 22.1844 seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-5

Perfect score: 1367

Sequence: 1 MLLAIDVNTHTVVVGLLSGM.....LRLVFNLEVQVGRGLKTAR 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1367	100.0	272	16	006282
2	1242	90.9	274	16	Q9CD56
3	698.5	51.1	265	16	Q9XN6
4	520.5	38.1	259	16	Q8XHL5
5	503.5	36.8	255	16	Q8R7M2
6	487.5	35.7	254	16	Q9KGH5
7	476.5	34.9	258	2	Q9F985
8	467.5	34.2	273	16	Q97EB4
9	456.5	33.4	259	16	Q8YAC5
10	448.5	32.8	259	16	Q92F54
11	404	29.6	262	16	Q9RX54
12	382.5	28.0	261	16	Q9A621
13	299.5	21.9	256	16	Q8RFE4
14	286	20.9	246	16	Q9WZT5
15	254	18.6	273	16	Q83446
16	209.5	15.3	212	2	O32514

17	152	11.1	592	16	Q9JW17
18	149	10.9	248	16	Q9HWC1
19	147.5	10.8	592	16	Q9JX21
20	138	10.1	262	16	O51477
21	115.5	8.4	295	16	Q8Y2M4
22	114	8.3	229	16	O67753
23	112	8.2	242	16	Q9PC14
24	108.5	7.9	356	2	Q9K506
25	108.5	7.9	478	2	Q9K505
26	108.5	7.9	487	2	Q9K504
27	106	7.8	257	16	P74045
28	101	7.4	732	17	O27676
29	100.5	7.4	264	16	O08343
30	100	7.3	255	16	Q928N2
31	100	7.3	276	16	Q8YQD7
32	98	7.2	268	2	Q9LCS2
33	97.5	7.1	223	16	Q9ZKY6
34	97.5	7.1	4735	2	O54666
35	96	7.0	730	16	Q9RU93
36	94.5	6.9	255	16	Q8Y4N4
37	93.5	6.8	746	16	Q9HZ69
38	93	6.8	405	10	Q9M2V1
39	92.5	6.8	465	17	Q9HSN0
40	92	6.7	462	5	Q9XWFO
41	92	6.7	649	2	Q9R6F3
42	92	6.7	1510	2	Q9Z465
43	91.5	6.7	209	16	Q9PIA9
44	91	6.7	391	16	Q9A7V5
45	91	6.7	449	16	Q9A3Q1

ALIGNMENTS

RESULT 1

O06282 ID O06282 PRELIMINARY; PRT: 272 AA.

AC O06282;

DT 01-JUL-1997 (TREMBlrel. 04, Created)

DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative, Baf family).

GN RV3600C OR MTCY07H7B.22 OR MT3706.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV.

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,

RA Badcock K., Basham D., Brown D., Brown T., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

complete genome sequence.";

RL Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / OSHKOSH;

RA Fleischmann J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains.";

```

DR EMBL; AE007092; AAK46859.1; -.
DR TIGR; MT2555; -.
DR Tuberculist; RV2482c; -.
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
KW Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
KW Complete proteome.
FT CONFLICT 180 180 R -> Q (IN REF. 2).
FT CONFLICT 778 778 C -> R (IN REF. 2).
SQ SEQUENCE 789 AA; 88314 MW; B78D7D8F9296AE22 CRC64;

      Query Match          6.4%; Score 87; DB 1; Length 789;
Best Local Similarity    24.4%; Pred. No. 18;
Matches 81; Conservative 27; Mismatches 120; Indels 104; Gaps 19

QY   2 LLAIDVRNTHVTVGLLSMKKEHAKVV---QQRWRTTSEVTVADELALTIDGLLGESERL 58
Db   :|: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 ILRTDPRRARVAG-----ESAKVSELQQWRDTTVAEHK-----RDFAQF 165

QY   59 TGTAALSTVPSLVHEVRIMLQYPSVPVHLIEPGVRTGPILLVDNPKEYGADRIVNCLA 118
Db   :|: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 166 VSRALLALARA--EYRILGPQY--KSPR-LVKP-----EMLASAREFRAGLDRI---- 209

QY   119 AYDFRKAAIVYDFG-----SSICVDVVS AKGEFLGGATAPGVQVSSDAARAARS 168
Db   :|: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 210 -----PCGATVEDACKMLDELTGSQSVSDLVSLVGLASRGDFPEFDYEYQVAAMRA 263

QY   169 ALR---RVELARPRAW--GKNITVECMQAGA-----VFEGFAGLVDLVGR-----IR 210
Db   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 264 ALEAHPAVLFFSHRSYIDGVVVPVQAQNRLPPVHMEGGINLSEGLMGLMRBSGMIFIR 323

QY   211 EDVS-----GFSDVH-----DVATVAT-GHTAPLLLPEL-----HTVDHY----- 244
Db   :|: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 324 RNIGNDPLYKYVLKEYGVYVVEKRFNLNSWSTEGTRSRTSKMLPKPKGLMSYVADAYLDGR 383

QY   245 DQHLLTQLGLRLVFERNLEV-----QRGLRKT 270
Db   :|: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 384 SDDILLQGVISCFDQLHEITEYAAYARGAETK 415

Search completed: June 24, 2003, 21:51:08
Job time : 5.73083 secs
```

Search completed: June 24, 2003, 21:51:08
Job time : 5.73083 secs

```
CC very different kinetic properties than the larger NADP-dependent
CC trifunctional enzyme and is unique in that it requires formation
CC of an enzyme-magnesium complex to allow binding of NAD.
CC
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: J04627; AAA39827.1; -.
CC EMBL: M63445; AAA39828.1; -.
CC EMBL: M63415; AAA39828.1; JOINED.
CC EMBL: M63439; AAA39828.1; JOINED.
CC EMBL: M63440; AAA39828.1; JOINED.
CC EMBL: M63441; AAA39828.1; JOINED.
CC EMBL: M63442; AAA39828.1; JOINED.
CC EMBL: M63443; AAA39828.1; JOINED.
CC EMBL: M63444; AAA39828.1; JOINED.
CC EMBL: M63445; AAA39828.1; JOINED.
CC EMBL: S52980; -. NOT_ANNOTATED_CDS.
CC PIR: A33267; A33267.
CC HSP: P11586; 1A4I.
CC
CC MGD; MGI:1338850; Mthfd2.
CC InterPro: IPR000672; THE_Dhg_Cyh.
CC Pfam: PF00763; THE_DHG_CYH_1.
CC Pfam: PF02882; THE_DHG_CYH_2; 1.
CC PRINTS: PR00095; THEHDRGNASE.
CC PRODOM: PD002300; THE_Dhg_Cyh_1.
CC PROSITE: PS00766; THE_DHG_CYH_1; 1.
CC PROSITE: PS00767; THE_DHG_CYH_2; 1.
CC MULTIFUNCTIONAL ENZYME: One-carbon metabolism; Oxidoreductase; NAD;
CC Hydrolase; Mitochondrion; Transit peptide; Magnesium.
CC TRANSIT 1 35 MITOCHONDRION.
CC CHAIN 36 350 BIFUNCTIONAL METHYLENETETRAHYDROFOLATE
CC DEHYDROGENASE/CYCLOHYDROLASE.
CC SEQUENCE 350 AA; 37863 MW; 896AD40D9154E9D7 CRC64;
CC
CC Query Match 6.4%; Score 87.5; DB 1; Length 350;
CC Best Local Similarity 20.7%; Pred No. 6.2;
CC Matches 52; Conservative 35; Mismatches 75; Indels 89; Gaps 10;
CC
CC 9 NHTVVGLLSGMKEHAKVQVQWRINTESEVTADELALITDGLIGE-----DSERLTGT 61
CC I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I:
CC 98 NSETIV-----KPAVSSEELLNSIRKLNDE---NVDGLLVQLPLPEHDERKVCNA 147
CC
CC 62 AALSTVPSVLHEV---RMLDQY--WPSVPHVLTEPGVRTGTGPIILVDN----- 104
CC I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I:
CC 148 VSPDKVDGDFHVINVGRCMLDQYSMLPAPGWVVEIKRTIPTLGRKNVVAGRSKNVGM 207
CC
CC 105 -----PKE-----VCADRIVNCLAAAYDRFR 124
CC I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I:
CC 208 PIAMLLHTDGAHERPGGDATVTISHRYTPKQELKHTTILADIVISAAGIPNLITA-DMIK 266
CC
CC 125 KAAIYVDFGSSICVDVWSAKGEFLG-----GAIAP---GVQVSSDAARAASAL 170
CC I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I:
CC 267 EGAAVIDVGINRVQDPVTAKEPLKLVGDVDFEGVKKKAGVITPVPGVGPMVTYMLMKNITII 326
CC
CC 171 RVELARPRSV 181
CC I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I:
CC 327 AAKVLRPEEL 337
CC
CC RESULT 14
CC METH_MYCLE
CC ID METH_MYCLE STANDARD; PRT; 1206 AA.
CC AC Q49775; Q9S378; Q9CC37;
CC DT 15-JUL-1998 (Rel. 36, Created)
```

```
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)
DE (Methionine synthase, vitamin-B12 dependent isozyme) (MS).
GN METH OR ML1307 OR MLCB2533.04 OR B2126_C1_157.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1169;
RN [1]
RN SEQUENCE FROM N.A.
RP Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- CATALYTIC ACTIVITY: 5-methyltetrahydrofolate + L-homocysteine -
CC tetrahydrofolate + L-methionine.
CC -1- COFACTOR: COBALAMIN (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC -1- SIMILARITY: BELONGS TO THE VITAMIN-B12 DEPENDENT METHIONINE
CC SYNTHASE FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 873.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: U00017; AAA17182.1; ALT_FRAME.
CC EMBL: AL035310; CAA22918.1; ALT_INIT.
CC EMBL: AL583921; CAC31688.1; -.
CC HSP: P13009; 1BMT.
CC Lepoma; ML1307; -.
CC InterPro: IPR003312; B12-binding.
CC InterPro: IPR003759; CoMet_synt_B12.
CC InterPro: IPR000489; Dhdropt_synt.
CC InterPro: IPR004223; Met_synt_B12.
CC InterPro: IPR003726; S_methyl_trans.
CC Pfam: PF00809; pterin_bind; 1.
CC Pfam: PF02310; B12-binding; 1.
CC Pfam: PF02574; S-methyl_trans; 1.
CC Pfam: PF02607; B12-binding_2; 1.
CC Pfam: PF02965; Met_synt_B12; 1.
CC Transferrase; Methyltransferase; Methionine biosynthesis; Vitamin B12;
CC Cobalt; Complete proteome.
CC DOMAIN 751 830 COBALAMIN-BINDING (POTENTIAL).
CC METAL 753 753 COBALT (POTENTIAL).
CC SEQUENCE 1206 AA; 132392 MW; 7786CE5307D7CA86 CRC64;
CC
CC Query Match 6.4%; Score 87.5; DB 1; Length 1206;
CC Best Local Similarity 23.0%; Pred No. 27;
CC Matches 61; Conservative 30; Mismatches 95; Indels 79; Gaps 14;
CC
CC 35 ESEVTADELALITDGLIGEDSERLTGTAALSTVPSVLHEVRLMDQYWP-SYP-----HVL 89
CC I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I: I:
CC 272 EYLPQDELAELAGFAEFGSLVVG-GCCGCTTPDHIREVAAVAARVNDGTVPGRHVT 330
CC DB
```


SEQUENCE FROM N.A. (ISOFORM TRAD).
 MEDLINE-98181067; PubMed-9512535;
 Cartwright R., Dunn A.M., Simpson P.J., Tambini C.E., Thacker J.;
 "Isolation of novel human and mouse genes of the recA/RAD51
 recombination-repair gene family.";
 Nucleic Acids Res. 26:1653-1659(1998).
 [2]
 SEQUENCE FROM N.A. (ISOFORM TRAD).
 MEDLINE-98234549; PubMed-9570954;
 Pittman D.L., Weinberg L.R., Schimenti J.C.;
 "Identification, characterization, and genetic mapping of Rad51, a
 new mouse and human RAD51/RecA-related gene.";
 Genomics 49:103-111(1998).
 [3]
 SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS TRAD-D1/D2 TO
 TRAD-D8).
 TISSUE=Brain;
 MEDLINE-99194570; PubMed-10092526;
 Kawabata M., Saeki K.;
 "Multiple alternative transcripts of the human homologue of the mouse
 TRAD/R51H3/RAD51D gene, a member of the recA/RAD51 gene family.";
 Biochem. Biophys. Res. Commun. 257:156-162(1999).
 CC -1- FUNCTION: MAY HAVE A ROLE IN DNA REPAIR AND IN MEIOTIC
 RECOMBINATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS: 7 ISOFORMS; TRAD (SHOWN HERE), TRAD-D1/D2,
 TRAD-D3, TRAD-D4, TRAD-D5, TRAD-D6/D7 AND TRAD-D8; ARE PRODUCED BY
 ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN COLON, PROSTATE, SPLEEN, TESTIS,
 OVARY, THYMUS AND SMALL INTESTINE. WEAKLY EXPRESSED IN LEUKOCYTES.
 CC -1- SIMILARITY: BELONGS TO THE RECA FAMILY. RAD51 SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y15572; CAA75681.1; -
 DR EMBL; AF034956; AAC39719.1; -
 DR EMBL; AB013341; BAA25914.1; -
 DR EMBL; AB016223; BAA31747.1; -
 DR EMBL; AB016224; BAA31748.1; -
 DR EMBL; AB016225; BAA31749.1; -
 DR EMBL; AB018360; BAA33779.1; -
 DR EMBL; AB018361; BAA33780.1; -
 DR EMBL; AB018362; BAA33781.1; -
 DR EMBL; AB018363; BAA33782.1; -
 DR EMBL; AB020412; BAA34690.1; -
 DR Genew; HGNC:9823; RAD51L3.
 DR MIM; 602954; -
 DR InterPro; IPR003265; Endo_3c.
 DR InterPro; IPR001553; RecA.
 DR Pfam; PF00730; HhH-GPD; 1.
 DR ProDom; PD000229; RecA; 2.
 DR PROSITE; PS0162; RECA; 2; 1.
 DR DNA-binding; ATP-binding; Nuclear protein; Alternative splicing.
 KW NP_BIND 107 114
 FT DOMAIN 200 205
 FT POLY-VAL.
 FT VARSPLIC 49 49
 FT MISSING (IN ISOFORM TRAD-D1/D2).
 FT VARSPLIC 50 328
 FT MISSING (IN ISOFORM TRAD-D3).
 FT VARSPLIC 116 160
 FT MISSING (IN ISOFORM TRAD-D4).
 FT SLDKLLDAGLTGCE -> RQKLSGGRWCHML (IN
 ISOFORM TRAD-D5).
 FT SLDKLLDAGLTGTEIVTEIVGGPSGKTQVCL -> RHGGRT
 QVTEWEDCSCLRSPQDGRGVSGML (IN ISOFORM
 TRAD-D6/D7).
 FT VARSPLIC 119 328
 FT MISSING (IN ISOFORM TRAD-D6/D7).
 FT VARSPLIC 193 212
 FT VTGSGTGKVVVDVSTAVV -> DGIPEHLNHPCLHVVH
 LFC (IN ISOFORM TRAD-D8).

FT VARSPLIC 213 328 MISSING (IN ISOFORM TRAD-D8).
 SQ SEQUENCE 328 AA; 35049 MW; 6038DA9356DF354A CRC64;
 Query Match 6.4%; Score 88; DB 1; Length 328;
 Best Local Similarity 22.0%; Pred. No. 5.2;
 Matches 55; Conservative 38; Mismatches 89; Indels 68; Gaps 10;
 QY 14 VGLLSGM-KEHAKVVOQWRIIT-----ESEYTADELALTIDGLI----- 51
 DB 6 VGLCPGLTEEMIQLLRSRIKTVVDLVSADEEVAQKGLSKALVALRRVLLAAQSAFP 65
 QY 52 --GED--SERLTCTAALSTVPSVLHEVRIMLDQYWPSPVHLIEPGVRTG-IPLLVDNPK 106
 DB 66 VNGADLYEELKTSTAILSTGIGSLDK-----LLDAGLYTGEIVTEIVGGPG 110
 QY 107 EVGADRIVNCLAAAYDRFRKAAIVVDFGSSICVDVWSAKGEFLGATAPGVQVSSDAAAAR 166
 DB 111 S-GKTQVCLCMAAN-----VAIGLQONVLYVDSNGGLTASRLQLLQAQTDEEQ 160
 QY 167 SAALRVELARPRSVVGNKTVCEQAGAVFGAGLVLDGLVGRIDEDVSGFSDVDHVAIV- 225
 DB 161 AEALRRIQV-----VHAFDIFOMLDVLQELRGTVAAQVTSSTGTVKVVVD 206
 QY 226 -ATGHTAPLL 234
 DB 207 SVTAVVSPLL 216
 RESULT 13
 MTDC_MOUSE STANDARD; PRT; 350 AA.
 ID MTDC_MOUSE
 AC P18155;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase,
 mitochondrial precursor [includes: NAD-dependent
 methylenetetrahydrofolate dehydrogenase (EC 1.5.1.15);
 methylenetetrahydrofolate cyclohydrolase (EC 3.5.4.9)].
 GN MTHFD2 OR NMDMC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 36-55.
 RX MEDLINE-89174757; PubMed-2647744;
 RA Belanger C., Mackenzie R.E.;
 "Isolation and characterization of cDNA clones encoding the murine
 NAD-dependent methylenetetrahydrofolate dehydrogenase-
 methylenetetrahydrofolate cyclohydrolase.";
 J. Biol. Chem. 264:4837-4843(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91133660; PubMed-1999293;
 RA Belanger C., Mackenzie R.E.;
 "Structural organization of the murine gene encoding NAD-dependent
 methylenetetrahydrofolate dehydrogenase-methylenetetrahydrofolate
 cyclohydrolase.";
 Gene 97:283-288(1991).
 RN [3]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE-91360330; PubMed-1843253;
 RA Belanger C., Peri K.G., Mackenzie R.E.;
 "Analysis of the promoter region of the gene encoding NAD-dependent
 methylenetetrahydrofolate dehydrogenase-methylenetetrahydrofolate
 cyclohydrolase.";
 Nucleic Acids Res. 19:4341-4345(1991).
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NAD(+) =
 5,10-methylenetetrahydrofolate + NADH.
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
 formyltetrahydrofolate.
 CC -1- COFACTOR: Magnesium. This NAD-dependent bifunctional enzyme has

Oxidoreductase; Nitrogen fixation; Electron transport; Iron-sulfur;
 4Fe-4S; Complete proteome.

KW 690 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
 KW METAL 690 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
 FT METAL 693 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
 FT METAL 696 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
 FT METAL 700 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
 FT METAL 746 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
 FT METAL 749 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
 FT METAL 752 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
 FT METAL 756 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).
 SQ SEQUENCE 1199 AA; 131457 MW; 12F5C34C3B8D50B6 CRC64;

Query Match 6.5%; Score 88.5; DB 1; Length 1199;
 Best Local Similarity 24.9%; Pred. No. 22;
 Matches 59; Conservative 33; Mismatches 86; Indels 59; Gaps 12;

QY 58 LGTAALSTVPSVLHEVRLMDQYWPSP-----HVLIEPGVGTGTLPLVDNPK 106
 DB 9 LDGNEAVARVAYLLSEV-IAIYPIPTSPMGSEWDAAAEH---RPNLNGTVPLVEMQS 64

Y 107 EYGDRIYNCLAAVDRFRKAAIIVDFGSSICVDVVSXKGEFLGGAIAPGVQVSSDAAAR 166
 DB 65 EGGAGTV-----HGAQSGALTTFITASQGLMLPWNHKTAGELTAMVL-----HVAAR 115

QY 167 SAALRRVELARPRS--VVGKNTVECMQAGAVFGFAGLVLDGLVGRIRDVSGFSVD--HDV 222
 DB 116 SLAAQGLSTFGDHSVMAARNT-----GFAML-----SSNSVQEAHDF 153

QY 223 AIATVGTHTAPLLPELHVVD-----HYDQHLTL-----QGLR-IVFERNLEVQGRUKT 270
 DB 154 ALIATATSPATRIPLGHFFDGPRTSHEEQKIELLPQEVLRGLIKEDVLAHGRALT 210

RESULT 11
 YQBC_BACSU
 ID YQBC_BACSU STANDARD; PRT; 297 AA.
 AC P54448;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein yqec.
 GN YQEC.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 ON NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 SC STRAIN=168 / JH642;
 TC MEDLINE=97124195; PubMed=8969508;
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes.";
 RL Microbiology 142:3103-3111(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S., Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruns C., Bruns C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C., Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";
 RL Nature 390:249-256(1997).
 CC -!- SIMILARITY: BELONGS TO THE 6-PHOSPHOGLUCONATE DEHYDROGENASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
 CC EMBL; D84432; BAAL2439.1; -;
 DR EMBL; Z99117; CAB14514.1; -;
 DR HSSP; P00349; 2PGD.
 DR Subtilist; BG11631; yqec.
 DR InterPro; IPR001744; 6PGD.
 DR InterPro; IPR004849; Gnd_rel.
 DR Pfam; PF00393; 6PGD; 1.
 DR Pfam; PF03446; NAD_binding_2; 1.
 DR TIGRFAMs; TIGR00872; gnd_rel; 1.
 DR PROSITE; PS00461; 6PGD; FALSE_NEG.
 KW Hypothetical protein; Oxidoreductase; Complete proteome.
 SQ SEQUENCE 297 AA; 32769 MW; 83607681471AC73C CRC64;

Query Match 6.4%; Score 88; DB 1; Length 297;
 Best Local Similarity 28.0%; Pred. No. 4.6;
 Matches 35; Conservative 20; Mismatches 60; Indels 10; Gaps 4;

QY 14 VLLSGMKEHAKVVOQWRTSEVTADELALTIDGLIGEDSERLTGTAALSTVPSVLHE 73
 DB 3 IGLIGLGRMGINIGKQFIDRNHVVGVYDVNQAAVDLKAAYGAE---GFTNLKEFISLLHP 59
 QY 74 VRIMLDQWPSVPHVIEPGVGTGTLPLV--DNPKEVGADRIVNCCLAAVDRFRKAAI-VV 130
 DB 60 PRIL-----WVWPHGIVDAVLDSVFLSKGDMITEAGNSHYKESIRRYNQMKAGIHYL 115

QY 131 DFGSS 135
 DB 116 DAGTS 120

RESULT 12
 R51D_HUMAN
 ID R51D_HUMAN STANDARD; PRT; 328 AA.
 AC 075771; 060355; 043537; 076085; 075196; 075847; 075848; 076073;
 AC 094908;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA repair protein RAD51 homolog 4 (R51H3) (TRAD).
 GN RAD51L3 OR RAD51D.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

DR Pfam: PF00430; ALAD; 1.
DR PRINTS: PR00144; DALDHYDRATASE.
DR PRODOM: PD002304; Alad_dehydratase; 1.
DR PROSITE: PS00169; D_ALA_DEHYDRATASE; 1.
DR Porphyrin biosynthesis; Lyase; Magnesium; Complete proteome.
FT ACT_SITE 260 260 BY SIMILARITY
SQ SEQUENCE 337 AA; 37037 MW; 7D650C49795D542C CRC64;

Query Match 6.5%; Score 89; DB 1; Length 337;
Best Local Similarity 22.1%; Pred. No. 4.5;
Matches 48; Conservative 35; Mismatches 70; Indels 64; Gaps 9;

QY 31 RIRTESVTVADELALTIDGLIGEDSERLTGTAALSTVPSVLUHEVRIMLDQWTWPSVPHVLI 90
DB 26 RLVRNRYLVDDLLPLFVLDGVNOR-----ESIPSPMGVERLSIDQ-----LLI 70
QY 91 --PEGVGTGPIVLVDRNPKVEGADRVNCLAAAY-----DRFKAATVVDVFGS 134
DB 71 EAEWVALGIPALALFFPTVPVEKSLDAAEAINEGIAQRAIIRERFPELGIITD--- 127
QY 135 SICVDVVSARGKEFGFLGAPGVQVSSDAAAARSAAALRRVELARPSRVGKNTVECMQAGA 194
DB 128 -VALDPPTTHGQ--DGILDDGGVYLVDSI-----DVLVRQALSHAEAGA 169
QY 195 -VFGFAGLVGLVGRIRREDVSGFSVDHDAIVATGHT 230
DB 170 QVWAPSDMDGGRIGAIIE-----ALESAGHT 195

RESULT 10
NIFU_SYNY3
ID NIFU_SYNY3 STANDARD; PRT; 1199 AA.
AC A52965;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative pyruvate-flavodoxin oxidoreductase (EC 1.2.7.-).
GN NIFU CR SLL0741.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiyama M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- FUNCTION: OXIDOREDUCTASE REQUIRED FOR THE TRANSFER OF ELECTRONS
CC FROM PYRUVATE TO FLAVODOXIN (POTENTIAL).
CC -1- CATALYTIC ACTIVITY: Pyruvate + CoA + oxidized flavodoxin = acetyl-
CC CoA + CO(2) + reduced flavodoxin.
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D64005; BAA10774.1; -
CC InterPro: IPR001450; 4Fe4S_ferredoxin.
CC InterPro: IPR002869; POR.
CC InterPro: IPR002880; POR_N.
CC Pfam: PF00037; fer4; 2.
CC Pfam: PF01558; POR; 1.
CC Pfam: FF01855; POR_N; 1.
CC PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.

RESULT 7

```

DHOM_SINY3
ID DHOM_SINY3 STANDARD; PRT; 433 AA.
AC P52986;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Homoserine dehydrogenase (EC 1.1.1.3) (HDH).
GN HOM OR SL10455.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=11148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Suglura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: L-homoserine + NAD(P)(+) -> L-aspartate 4-
CC semialdehyde + NAD(P)H.
CC -1- PATHWAY: THIRD STEP IN CONVERSION OF L-ASPARTATE TO HOMOSERINE.
CC HOMOSERINE PARTICIPATES IN THE BIOSYNTHESIS OF THREONINE AND THEN
CC ISOLEUCINE AND IN THE BIOSYNTHESIS OF METHIONINE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HOMOSERINE DEHYDROGENASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D64001; BAA10325.1;
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001342; Homoserine_dh.
DR InterPro; IPR005106; NAD_binding_3.
DR Pfam; PF00742; Homoserine_dh; 1.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF03447; NAD_binding_3; 1.
DR PROSITE; PS01042; HOMOSER.DHGENASE; 1.
DR Oxidoreductase; NADP; Threonine biosynthesis; Isoleucine biosynthesis;
KW Methionine biosynthesis; Complete proteome.
NT NP_BIND 5 12 NADP (POTENTIAL).
SEQUENCE 433 AA; 45486 MW; 4A7A034A7D2766B5 CRC64;
-----
Query Match 6.7%; Score 92; DB 1; Length 433;
Best Local Similarity 23.2%; Pred. No. 3.5;
Matches 60; Conservative 45; Mismatches 110; Indels 44; Gaps 9;
QY 1 MLLAIDVNRHTVVGLLSGMKEHAKVVOQMRITSE--VVADELALTDIGLIGEDSERL 58
DB :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
30 LLKAEVK---AVGVSRLDPRQVNLPPPELVTTDLAIVTDPDIAIWMELMGLEPARS 85
QY 59 TGTAAISTVPSVLHEVRIMLDQYPSV-----PHVLEPGVRTGIPLVDNPKVEGA 110
DB :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
86 LILQAIHAKHIVTANKAVIARYGPEIYEANQHGVTYLLLEAAVGGGIPKIKLQSLGG 145
QY 111 DR-----IVNCLAAVDRFRKAAIWDVDFGSSICVDVYSKAGFEGLGGAIA-PGVQVSSDAA 163
DB :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
146 NRIOSIVGILNGTNYILSRMTSEGADF-----DEVLTAAQLGCGYAEADPSADVVDGLDA 199
QY 164 AARSAAL-----RRVELARP--RSVVGKNTVECMQACAGVFAGFLVDGLVG----- 207
DB :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
200 ADKATAILASLFGCGRVKREDVACBGIRSVSAVDIAYADRLGFLVKLLAIADGNAGEDSEA 259
QY 208 -RREDVSGFSDHDAIV 225
DB :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
260 LQLRVHPTLIAKDHPLASV 278

```

RESULT 8

```

AMPN_STRLI
ID AMPN_STRLI STANDARD; PRT; 857 AA.
AC Q11010;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aminopeptidase N (EC 3.4.11.2) (Lysyl aminopeptidase) (Lys-AP)
DE (Alanine aminopeptidase).
GN PEPN.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 415-439.
RX STRAIN=66 / 1326;
RX MEDLINE=94215872; PubMed=7909302;
RA Butler M.J., Aphale J.S., Binnie C., Dizonno M.A., Krygsmann P.,
RA Soltis G.A., Walczyk E., Malek L.T.;
RT "The aminopeptidase N-encoding pepN gene of Streptomyces lividans
RT 66.";
RL Gene 141:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=66 / 1326;
RX MEDLINE=94176084; PubMed=7765336;
RA Butler M.J., Aphale J.S., Dizonno M.A., Krygsmann P., Walczyk E.,
RA Malek L.T.;
RT "Intracellular aminopeptidases in Streptomyces lividans 66.";
RL J. Ind. Microbiol. 13:24-29(1994).
CC -1- FUNCTION: AMINOPEPTIDASE WITH BROAD SUBSTRATE SPECIFICITY TO
CC SEVERAL PEPTIDES. SHOWS STRONG PREFERENCE FOR LEUCINE BUT CLEAVES
CC ALSO NEXT TO ARG AND LYSINE IN PEPTIDE-BOND-CONTAINING SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-1-
CC Xbb- from a peptide, amide or arylamide. Xaa is preferably Ala,
CC but may be most amino acids including Pro (slow action). When a
CC terminal hydrophobic residue is followed by a prolyl residue, the
CC two may be released as an intact Xaa-Pro dipeptide.
CC -1- COFACTOR: BINDS 1 ZINC ION.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PFM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L23172; AAA26696.1;
DR MEROPS; M01.009;
DR InterPro; IPR001930; Ala_peptase.
DR InterPro; IPR000130; Zn_Mtppeptidse.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc.
FT METAL 298 298 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 299 299 BY SIMILARITY.
FT METAL 302 302 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 399 399 PROTON DONOR (POTENTIAL).
SQ SEQUENCE 857 AA; 94423 MW; 40B749F78D201A7D CRC64;
-----
Query Match 6.5%; Score 89.5; DB 1; Length 857;
Best Local Similarity 23.5%; Pred. No. 12;
Matches 47; Conservative 31; Mismatches 79; Indels 43; Gaps 9;

```

```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN-BP504;
RX  MEDLINE=93325323; PubMed=7601846;
RA  Dashazer D., Wood G.E., Friedman R.L.;
RT  "Identification of a Bordetella pertussis regulatory factor required
RT  for transcription of the pertussis toxin operon in Escherichia
RL  coli.";
RN  J. Bacteriol. 177:3801-3807(1995).
[2]
RN  SEQUENCE OF 1-38 FROM N.A.
RP  STRAIN-BP504;
RA  Wood G.E., Friedman R.L.;
RT  "Identification of a birA homolog in Bordetella pertussis.";
RL  Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN  SEQUENCE OF 239-267 FROM N.A.
RP  STRAIN-BP536;
RX  MEDLINE=96419162; PubMed=8821935;
RA  Allen A.G., Maskell D.J.;
RT  "The identification, cloning and mutagenesis of a genetic locus
RT  required for lipopolysaccharide biosynthesis in Bordetella
RL  pertussis.";
RN  Mol. Microbiol. 19:37-52(1996).
-1- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC  A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC  RNA POLYMERASE.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; U12020; AAA75361.1; -
DR  EMBL; AF016461; AAC68834.1; -
DR  EMBL; X90711; CAA62242.1; -
DR  InterPro; IPR004619; Baf.
DR  Pfam; PF03309; Bvq_acc_factor; 1.
DR  Transcription regulation; Activator.
KW  SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;
Query Match 6.8%; Score 92.5; DB 1; Length 267;
Best Local Similarity 28.6%; Pred. No. 1.8;
Matches 36; Conservative 16; Mismatches 57; Indels 17; Gaps 5;
-----
93 GVTGTPLLVDNPKVEGADR---IVNCLAAAYDRFKAIAIVVDGSSICVDVVSAGKEFLG 149
91 GLRNG-----YRNPDLGADRWACMGVGLARQPSVHPPLLVASFATTLDTIGPDNVFPG 146
150 GATAPGVQVSSDAAARSALRVELARPRSVVGNKNTVECMQA---GAVGFEAGLV--DG 204
147 GLILPGAMRGALAYGTAHLPLAD-----GLVADYPIDTHQAIASGIAAQAIVROW 201
205 LVGRIR 210
202 LAGRQR 207
RESULT 6
CH60_MYCGE
ID  CH60_MYCGE STANDARD; PRT; 543 AA.
AC  P47632; Q49358;
DT  01-FEB-1996 (Rel. 33, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DE  16-OCT-2001 (Rel. 40, Last annotation update)
DE  60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN  GROEL OR MOPA OR GROEL OR MG392.
OS  Mycoplasma genitalium.
OC  Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX  NCBI_TaxID=2097;

```

```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN-ATCC 33530 / G-37;
RX  MEDLINE=96026346; PubMed=7569993;
RA  Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA  Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA  Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA  Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA  Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA  Peterson S.N., Smith H.O., Hutchison C.A. III; Venter J.C.;
RT  "The minimal gene complement of Mycoplasma genitalium.";
RL  Science 270:397-403(1995).
[2]
RN  SEQUENCE OF 1-60 AND 466-543 FROM N.A.
RP  STRAIN-ATCC 33530 / G-37;
RX  MEDLINE=94075230; PubMed=8253680;
RA  Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT  "A survey of the Mycoplasma genitalium genome by using random
RT  sequencing.";
RL  J. Bacteriol. 175:7918-7930(1993).
-1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC  PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC  CONDITIONS (BY SIMILARITY).
CC  -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC  7 SUBUNITS (BY SIMILARITY).
CC  -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; U39721; AAC71620.1; -
DR  EMBL; U02252; AAD12515.1; ALT_INIT.
DR  EMBL; U02268; AAD12534.1; -
DR  HSSP; P06139; 1GLR.
DR  TIGR; MG392; -
DR  InterPro; IPR001844; Chaperin_Cpn60.
DR  InterPro; IPR002423; Cpn60_TCP-1.
DR  Pfam; PF00118; cpn60_TCP1; 1.
DR  PRINTS; PR00298; CHAPERONIN60.
DR  PRINTS; PR00304; TCOMPLEXTCP1.
DR  PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR  Chaperone; ATP-binding; Complete proteome.
KW  SEQUENCE 543 AA; 58354 MW; 80FA5C7037C4BA88 CRC64;
Query Match 6.8%; Score 92.5; DB 1; Length 543;
Best Local Similarity 20.9%; Pred. No. 4.1;
Matches 58; Conservative 43; Mismatches 92; Indels 85; Gaps 13;
-----
9Y 13 VVGLLSGMEKHAHVQOQR-----IRTESEVTADELALTIDGLIGEDSERLTGTAAL 64
Db 146 VAAISSGSEIGKLIQAAMALVGKNGVITDDAKTINTTLETTEGI-----EFKGTYS 199
9Y 55 STVPSVLHEVRIMLDQWPSVPHVLI-----EPGVTGIPLLVDNPKVEG 109
Db 200 PYWSDQEKMEVVLEQ-----PKILVSSLKINTIKELIPLEGSEVGNPLLVADP--F 252
9Y 110 ADRIVNCLAAAYDRFKAIAIVVDGSSICVDVVSAGKEFLG 155
Db 253 AEEVVTTLAV-NKLRGT-----INVAVKCNVEYGERQKAALDIAISSGTLAYN 300
9Y 156 VQVSS---DAAARSALRVELARPRSVV--GNKNTVECMQAIVGFGAGLVGRIR 210
Db 301 TEINSGFKDVTVDNLGDARKVQIAKGKTTVIGGKGNKDKIKKH-----VELNGRLK 352
9Y 211 EDVSGFSVD-----HDVAIVATGHTAPILLPEL 238
Db 353 QTTDKYSDSLIKERIAIYLSOGVAVIRVGGATELAQKEL 390

```

```

OX NCBI_TaxID=2234;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RC MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kirschum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., Mckenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Gloeck A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001087; AAB90971.1; -.
CC TIGR; AF0267; -.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW Hypothetical protein; 23 POTENTIAL.
FT TRANSMEM 4
FT TRANSMEM 209 231 POTENTIAL.
FT SEQUENCE 597 AA; 66250 MW; E2BCA81C0E3B2540 CRC64;
CC -----
CC Query Match 7.1%; Score 96.5; DB 1; Length 597;
CC Best Local Similarity 19.3%; Pred. No. 2.2;
CC Matches 53; Conservative 41; Mismatches 81; Indels 99; Gaps 12;
CC -----
QY 41 DELALTIDGLIGEDSERLTGTAALSTV---PSVLHEVRIMLDQY-----W 82
DB 158 DKGACVACDGFVAMDDATSDGRVLMGRSFMFNPEVFHEVALLTEQYDGRGHRFVSVAPGF 217
QY 83 PSVPHVLIPEGVRTG---IPLLVDNPKVGGADRVNCLAAAYDRFRKAAIIVDFGSSI--C 137
DB 218 VGVTAAMSAGTAIGNDMVPAMDTRKFSVG---MCCLLT-----ARVVQVADLSDA 267.
QY 138 VDVV-----SAKGEFLGGAIPGVQVSSDAAAARSAAALRRVELAR-----PR 179
DB 268 VNMVKSGRGVPWLYIVGDKGREKGAV---LEVSADKFAVRYMDRYPEWAESELDFFPK 324
QY 180 SVVGKNTVCEMGAGAVFGAGLVGLVGRIREVDVSGFSVDHVAIVATGHTAPLLPELH 239
DB 325 QIEDKD-----DLVVVYVANH-----IVPEMY 345
QY 240 TVDHYD-----OHLTFLOGLRLVFERNLEVQGR 267
DB 345 SIVSYAVKDSLWRVYETLTGLILDSYGSIDVEXGK 379
CC -----
RESULT 5
BAF_BORPE
ID BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DT Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OC NCBI_TaxID=520;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.73083 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453a-5

Perfect score: 1367

Sequence: 1 MLLAIDVRNTHVTVVGLLSGM.....LRLVFNLEVQGRKLTAR 272

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	412	30.1	233	1	YACB_BACSU
2	108.5	7.9	543	1	CH60_MYCPN
3	97	7.1	473	1	FLI1_AGRF5
4	96.5	7.1	597	1	Y267_ARCFU
5	92.5	6.8	267	1	BAF_BORPE
6	92.5	6.8	543	1	CH60_MYCCE
7	92	6.7	433	1	DHOM_SYNY3
8	89.5	6.5	857	1	AMPN_STRLI
9	89	6.5	337	1	HEM2_PSEAE
10	88.5	6.5	1199	1	NTEJ_SYNY3
11	88	6.4	297	1	YQEC_BACSU
12	88	6.4	328	1	R51D_HUMAN
13	87.5	6.4	350	1	MTDC_MOUSE
14	87.5	6.4	1206	1	METH_MYCLE
15	87	6.4	789	1	PLSB_MYCTU
16	86.5	6.3	399	1	DP3B_MYCLE
17	86.5	6.3	461	1	PICC_RHOCA
18	86.5	6.3	598	1	IF2P_METRA
19	86	6.3	439	1	PURK_MYCLE
20	86	6.3	486	1	MURD_MYCTU
21	85.5	6.3	284	1	LPXD_METCA
22	85.5	6.3	467	1	FLI1_RHIME
23	85	6.2	311	1	PNK_MYCLE
24	85	6.2	686	1	CHEA_RHOSH
25	85	6.2	949	1	ALB1_RAT
26	84	6.1	285	1	PURR_BACSU
27	84	6.1	329	1	HEM2_MYCLE
28	84	6.1	777	1	PRZE_MYXHA
29	84	6.1	3172	1	PRY3_SACER
30	83.5	6.1	191	1	V645_METJA
31	83.5	6.1	323	1	SECF_ECOLI
32	83.5	6.1	344	1	MTDC_HUMAN
33	83.5	6.1	490	1	MURD_MYCLE

34 83.5 6.1 2269 1 RRPL_SV41
35 83 6.1 234 1 UREL_HELHE
36 83 6.1 310 1 ISPH_CHLPN
37 83 6.1 439 1 YF55_STRFR
38 83 6.1 1101 1 GUNC_CELFI
39 83 6.1 1550 1 GLTB_SYNY3
40 82.5 6.0 638 1 DXS_MYCTU
41 82.5 6.0 707 1 PNP_BUCAI
42 82.5 6.0 1596 1 ACS2_ACXY
43 82 6.0 267 1 HIS6_MYCTU
44 82 6.0 344 1 COA2_POVJC
45 81.5 6.0 148 1 TRBH_RHISN

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Tateuchi M., Tanakoshi A., Tanaka T., Terstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

QY 1 MLLAIDVRNTHTVVGLSGMKHAKVVOQWRIRTESEVTADALATIDGLIGDS---ER 57
Db 18 VILVLDVGNINVLGIYN---DTKLTAEWRLSDVLRSADEYGIQVNNLFQDOKLDP 73
QY 58 LTGTAALSTVPSVLHEVRIMLDQYWPSPVPHVLIPEGVGTGIPLLVDNPKVGGADRIVNC 117
Db 74 VEGTIISSVVPQSIENLRLNLSRYF-NVEPLVIGENAKLIDVRIEKPSEACADRLVNAI 132
QY 118 AAYDRFRKAAIIVDFGSSICVDVSAKGEFLGGAITAPGVQVSSDAAAARSALRRVELAR 177
Db 133 AAHEIYKRSLLIIDGTATTCFVRENGDYLGAICPGIKVSSEALFEKAAKLPVELIK 192
QY 178 PRSVGKNTVECMQAGVGFAGLVDGLVGRIRREDSVDFDHDVAIVATGHTAPLL 234
Db 193 PAYAICKNTISSIQGVIVRYLRQVKYLFELKLENLPDGRRTTSILVLATGGLAKLI 249

RESULT 14

US-09-813-453A-51
Sequence 51, Application US/09813453A
Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OGZ-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIORITY FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 51

LENGTH: 260

TYPE: PR

ORGANISM: Caulobacter crescentus

US-09-813-453A-51

Query Match 28.0%; Score 382.5; DB 9; Length 260;
Best Local Similarity 34.0%; Pred. No. 1.4e-26;
Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;

QY 1 MLLAIDVRNTHTVVGLSGMKHAKVVOQWRIRTESEVTADALATIDGLI---GEDSER 57
Db 1 MLLAIEQGNNTMFAIHG---ASVQAQWRSATSTRTADEYVVLVLSQSLSMQGLGFRA 56
QY 58 LTGTAALSTVPSVLHEVRIMLDQYWPSPVPHVLIPEGVGTGIPLLVDNPKVGGADRIVNC 117
Db 57 IDAVIISVVPQSIENLRLNLSRYF-NVEPLVIGENAKLIDVRIEKPSEACADRLVNAI 115
QY 118 AAYDRFRKAAIIVDFGSSICVDVSAKGEFLGGAITAPGVQVSSDAAAARSALRRVELAR 177
Db 116 GAAMYVPGVPLVVIDSGTATTFDVAADGAFEGGIAPGINLSMQALHEAAAKLPRIATQ 175
QY 178 P--RSVVGKNTVECMQAGVGFAGLVDGLVGRIRREDSVDFDHDVAIVATGHTAPLL 235
Db 176 PAGNRIVGDTVSAQSGVFWGIIISLIEGLVARIKAEGR-----EPMTVIATGGVASLFE 230
QY 236 PELHTVDHYDHLTQGLRLVPERN 260
Db 231 GATDSIDHFDSDLTIRGLLEIYRN 255

RESULT 15

US-09-813-453A-57
Sequence 57, Application US/09813453A
Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OGZ-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIORITY FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 57
LENGTH: 219
TYPE: PR
ORGANISM: Dehalococcoides ethenogenes
US-09-813-453A-57

Query Match 27.1%; Score 371; DB 9; Length 219;
Best Local Similarity 36.3%; Pred. No. 1.2e-25;
Matches 78; Conservative 49; Mismatches 80; Indels 8; Gaps 3;
QY 2 LLAIDVRNTHTVVGLSGMKHAKVVOQWRIRTESEVTADALATIDGLI---GEDSERL 58
Db 5 LVAVDIGNTSVNIIGIFEG---EKLANNHLSGVAQRMADYASLLGLLQHGAIHPEEL 60
QY 59 TGTAAALSTVPSVLHEVRIMLDQYWPSPVPHVLIPEGVGTGIPLLVDNPKVGGADRIVNC 118
Db 61 NRVIMCSVVPPLTTTFEEVFKSYFAAP-LVVGAGIKSGVKVMDNPREVGADRIVNA 119
QY 119 AYDRFRKAAIIVDFGSSICVDVSAKGEFLGGAITAPGVQVSSDAAAARSALRRVELAR 178
Db 120 ARVLYPGACIIVDMGTATTFDLSGGAYIGGAIAPGIATSAQAIKTSKLPKIEIRP 179
QY 179 RSVVGKNTVECMQAGVGFAGLVDGLVGRIRREDSV 213
Db 180 AKVIGSNTVSAMQSGIYFGIIGLVELVRRIOTEL 214

Search completed: June 24, 2003, 22:23:52
Job time: 12.4012 secs

```
Query Match      .31.6%; Score 431.5; DB 9; Length 258;
Best Local Similarity 37.6%; Pred. No. 5.5e-31;
Matches 100; Conservative 46; Mismatches 101; Indels 19; Gaps 4;

QY 1 MLLAIDVNTHTVTVGLSGMKEHAKVVOQWRIRTESEVTADELALTDIGLIGED 54
DB 1 MLLCIDCGTNTVTSVMDG----TDFATWRIATDHRRTADEYFVWNLTLMLQKLGLOGRI 56
QY 55 SERLTGTAAALSTVPSVLHEVRIMLDQYWPSPVPHVLEPGVGTGIPLLVDNPKVCADRIV 114
DB 57 SEALIS-----STARVFNFLVCLNRYFDCRPVYVVRKPGCELPVAPRVDPTTGVGPDRLV 112
QY 115 NCLAAIDFRKAAIIVDFGSSICVDVVSAGKEFLGGAIPAGVQVSSDAAAARSALRRVE 174
DB 113 NTVAGYDRHGDLIVDFGTATTEDVVPADPGAYIGVYAPCVNLSLEALHMAAALPHVD 172
QY 175 LARPSVVGKNTVCEMOMAGAVGAGLVGDLVGRIREVDSGFSVDHDAIVATGHTAPLL 234
DB 173 VTKPGVIGTNTVACIQSGVYWGVLGVEGIVRQIR-----MERDRPMKVATGGLASLF 227
235 LPELHTVDYDQHLTLQGLRLVFNERN 260
228 DLGFDLFDKVEDDLMHGRLIFDIN 253

RESULT 11
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-17

Query Match      30.1%; Score 412; DB 9; Length 233;
Best Local Similarity 39.4%; Pred. No. 2.7e-29;
Matches 84; Conservative 45; Mismatches 76; Indels 8; Gaps 3;

QY 1 MLLAIDVNTHTVTVGLSGMKEHAKVVOQWRIRTESEVTADELALTDIGLIGEDS---ER 57
DB 1 MLLVIDVNTGNTVLGVY----HDKLEVHWRIETSRHKTDEFEFGMLRSLFDHSLGMEQ 56
QY 58 LTGTAALSTVPSVLHEVRIMLDQYWPSPVPHVLEPGVGTGIPLLVDNPKVCADRIVNCL 117
DB 57 IDGIISSVWPPIMFALERMCTKYFTHBPQI-VGPGMKTGLNIRKYNPKVCADRIVNAV 115
QY 118 AAYDRFRKAAIIVDFGSSICVDVVSAGKEFLGGAIPAGVQVSSDAAAARSALRRVELAR 177
DB 116 AAHLXGNPLIVDFGATTTCYIDENKQYMGGAAPGCIITISTEALYSRAAKLPRIETR 175
QY 178 PRSVVGKNTVCEMOMAGAVGAGLVGDLVGRIR 210
DB 176 PDNIIGNKTVSAMQSGILFGVGVGVEGIVKRMK 208

RESULT 12
US-09-813-453A-8
; Sequence 8, Application US/09813453A
```

```
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Deinococcus radiopugnans
US-09-813-453A-8

Query Match      29.6%; Score 404; DB 9; Length 262;
Best Local Similarity 38.3%; Pred. No. 1.6e-28;
Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;

QY 2 LLAIDVRNTHTVVGL--LSGMKEHAKVVOQWRIRTESEVTADELALTDIGLIGEDSERLT 59
DB 6 LLAVDIGNTTTVLGLADASGALTHT-----WRIRTNREMLPDDLAQLHGLFTLAGAPIP 60
QY 60 GTAALSTV-PSVLHEVRIMLDQYW-----PSVPHVLEPGVGTGIPLLVDNPKVEG 109
DB 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDTPGSGV 110
QY 110 ADRIVNCLAA--YDRFRKAAIIVDFGSSICVDVVSAGKEFLGGAIPAGVQVSSDAAAARS 167
DB 111 ADRLCNLFGAEKYLGGLDYAVVDFGTSTNFDVWGRGRFLGGILATGAQVSADALFARA 170
QY 168 AALRRVELARPSVVGKNTVCEMOMAGAVGAGLVGDLVGRIREVDSGFSVDHDAIVAT 227
DB 171 AKLPRITLQAPETAIGKNTVHALQSGLVFGVAEWVGLLRRLRAELPG-----EAVAVAT 225
QY 228 GHTAPLLPELHTVDYDQHLTLQGLRLVFNERNLEVQ 264
DB 226 GGFSTVQGIQCEIDYDETLTLRGLVELWASRSEVR 262

RESULT 13
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

Query Match      28.5%; Score 389; DB 9; Length 250;
Best Local Similarity 34.2%; Pred. No. 3.4e-27;
Matches 81; Conservative 63; Mismatches 85; Indels 8; Gaps 3;
```


Db 181 VVGKNTVECMQAGVFGAGLVGRLVIRREDSVSGFSDHVAIVATGHTAPLLPELHT 240
QY 241 VDHYDQHLTLOGLRLVFNERNLEVRGLKTKAR 272
Db 241 VDHYDQHLTLOGLRLVFNERNLEVRGLKTKAR 272

RESULT 2

US-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 276
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-276

Query Match 99.7%; Score 1363; DB 9; Length 272;
Best Local Similarity 99.6%; Pred. No. 2.1e-114; Indels 0; Gaps 0;
Matches 271; Conservative 1; Mismatches 0;
QY 1 MLLAIDVRNTHVTVGLLSGKHEKAVVQWRIKRTSEVTADLALTDGLIGEDSERLTG 60
Db 1 VLLAIDVRNTHVTVGLLSGKHEKAVVQWRIKRTSEVTADLALTDGLIGEDSERLTG 60
QY 61 TAALSTVPSVLHEVRIMLDQYWPSPHVLIEPGVTRTGTPLLVNDKPEVGADRIVNCALAY 120
Db 61 TAALSTVPSVLHEVRIMLDQYWPSPHVLIEPGVTRTGTPLLVNDKPEVGADRIVNCALAY 120
QY 121 DRFRKAAIYVDFGSSICVDVVSAGKEFLGGATAPGVQVSSDAAARSALRVELARPRS 180
Db 121 DRFRKAAIYVDFGSSICVDVVSAGKEFLGGATAPGVQVSSDAAARSALRVELARPRS 180
QY 181 VVGKNTVECMQAGVFGAGLVGRLVIRREDSVSGFSDHVAIVATGHTAPLLPELHT 240
Db 181 VVGKNTVECMQAGVFGAGLVGRLVIRREDSVSGFSDHVAIVATGHTAPLLPELHT 240
QY 241 VDHYDQHLTLOGLRLVFNERNLEVRGLKTKAR 272
Db 241 VDHYDQHLTLOGLRLVFNERNLEVRGLKTKAR 272

RESULT 3

US-09-813-453A-4

Query Match 37.5%; Score 513; DB 9; Length 256;
Best Local Similarity 38.9%; Pred. No. 2.7e-38;

; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Vocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

Query Match 51.1%; Score 698.5; DB 9; Length 265;
Best Local Similarity 51.5%; Pred. No. 6.9e-55;
Matches 138; Conservative 53; Mismatches 64; Indels 13; Gaps 3;
QY 1 MLLAIDVRNTHVTVGLLSGKHEKAVVQWRIKRTSEVTADLALTDGLIGEDSERLTG 53
Db 1 MLLAIDVRNTHVTVGLLSGKHEKAVVQWRIKRTSEVTADLALTDGLIGEDSERLTG 53
QY 54 DSERLTGTAAALSTVPSVLHEVRIMLDQYWPSPHVLIEPGVTRTGTPLLVNDKPEVGADRI 113
Db 57 LGDGDIGTAICATVPSVLHEVRIMLDQYWPSPHVLIEPGVTRTGTPLLVNDKPEVGADRI 116
QY 114 VNCALAYDRFRKAAIYVDFGSSICVDVVSAGKEFLGGATAPGVQVSSDAAARSALRVR 173
Db 117 INAAVELYGGPAIVVDFGTATTFDASARGEYIGGVATGAPGIEISVEALGVKGALRKI 176
QY 174 ELARPRSIVGKNTVECMQAGVFGAGLVGRLVIRREDSVSGFSDHVAIVATGHTAPL 233
Db 177 EVARPRSIVGKNTVECMQAGVFGAGLVGRLVIRREDSVSGFSDHVAIVATGHTAPL 234
QY 234 LLPDLHTVDHYDQHLTLOGLRLVFNERNL 261
Db 235 VLGESSVIDEHEPFWLTLMGLRLVYERNV 262

RESULT 4

US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Vocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 12.4012 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-5
Perfect score: 1367
Sequence: 1 MLLAIDVRNTHVTVGLLSGM.....LRLVFNLEVRQGRKLTAR 272

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1367	100.0	272	9	US-09-813-453A-5
2	1363	99.7	272	9	US-09-712-363-276
3	698.5	51.1	265	9	US-09-813-453A-4
4	513	37.5	256	9	US-09-813-453A-55
5	509.5	37.3	255	9	US-09-813-453A-7
6	494.5	36.2	258	9	US-09-813-453A-2
7	487.5	35.7	254	9	US-09-813-453A-47
8	479.5	35.1	258	9	US-09-813-453A-49
9	471.5	34.5	262	9	US-09-813-453A-45
10	431.5	31.6	258	9	US-09-813-453A-6
11	412	30.1	233	9	US-09-813-453A-17
12	404	29.6	262	9	US-09-813-453A-8
13	389	28.5	250	9	US-09-813-453A-3
14	382.5	28.0	260	9	US-09-813-453A-51
15	371	27.1	219	9	US-09-813-453A-57
16	286	20.9	246	9	US-09-813-453A-9
17	274.5	20.1	257	9	US-09-813-453A-53
18	254	18.6	273	9	US-09-813-453A-10
19	209.5	15.3	212	9	US-09-813-453A-59

20	160.5	11.7	249	9	US-09-813-453A-61	Sequence 61, Appl
21	159	11.6	241	9	US-09-813-453A-63	Sequence 63, Appl
22	153	11.2	244	9	US-09-813-453A-41	Sequence 41, Appl
23	152	11.1	592	9	US-09-813-453A-22	Sequence 22, Appl
24	150.5	11.0	249	9	US-09-813-453A-70	Sequence 70, Appl
25	149	10.9	248	9	US-09-813-453A-20	Sequence 20, Appl
26	147.5	10.8	592	9	US-09-813-453A-43	Sequence 43, Appl
27	146.5	10.7	460	9	US-09-813-453A-39	Sequence 39, Appl
28	138	10.1	262	9	US-09-813-453A-11	Sequence 11, Appl
29	115	8.4	328	9	US-09-975-719-47	Sequence 47, Appl
30	114	8.3	229	9	US-09-813-453A-12	Sequence 12, Appl
31	112	8.2	242	9	US-09-813-453A-65	Sequence 65, Appl
32	106	7.8	257	9	US-09-813-453A-13	Sequence 13, Appl
33	100.5	7.4	264	9	US-09-712-363-183	Sequence 183, Appl
34	94	6.9	317	12	US-10-043-238-1	Sequence 1, Appl
35	94	6.9	317	12	US-10-043-238-3	Sequence 3, Appl
36	92.5	6.8	267	9	US-09-813-453A-15	Sequence 15, Appl
37	92	6.7	1510	9	US-09-738-626-3707	Sequence 3707, Appl
38	91.5	6.7	209	9	US-09-813-453A-21	Sequence 21, Appl
39	89.5	6.5	223	9	US-09-895-913A-74	Sequence 74, Appl
40	89.5	6.5	223	9	US-09-813-453A-14	Sequence 14, Appl
41	89.5	6.5	223	9	US-09-813-453A-67	Sequence 67, Appl
42	89	6.5	326	10	US-09-905-176-9	Sequence 9, Appl
43	89	6.5	337	10	US-09-815-242-12096	Sequence 12096, A
44	88	6.4	328	9	US-10-164-433-4	Sequence 4, Appl
45	87	6.4	789	9	US-09-712-363-244	Sequence 244, Appl

ALIGNMENTS

RESULT 1
US-09-813-453A-5
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

Query Match	100.0%	Score 1367;	DB 9;	Length 272;
Best Local Similarity	100.0%	Pred. No. 9.1e-115;		
Matches 272;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MLLAIDVRNTHVTVGLLSGMKEHAKVQOQWIRTESEVTADELALTIDGLIGEDSERLTG	60	
Db	1	MLLAIDVRNTHVTVGLLSGMKEHAKVQOQWIRTESEVTADELALTIDGLIGEDSERLTG	60	
Qy	61	TAALSTVPSVLHVEYRIMLDQYWPSPVPHVLPDQVGTGTPILVDNPKVGDRIYVNCCLAA	120	
Db	61	TAALSTVPSVLHVEYRIMLDQYWPSPVPHVLPDQVGTGTPILVDNPKVGDRIYVNCCLAA	120	
Qy	121	DRFRKAAIVDFGSSICVDVVSARGFEGAGVGVSSDAAARSAAARVELARPRS	180	
Db	121	DRFRKAAIVDFGSSICVDVVSARGFEGAGVGVSSDAAARSAAARVELARPRS	180	
Qy	181	VYGRNTVECMOAGAVFGVAGLVGDRIRVEDVSGFSVDHVDVAIVATGHTAPLLPELHT	240	
Db	181	VYGRNTVECMOAGAVFGVAGLVGDRIRVEDVSGFSVDHVDVAIVATGHTAPLLPELHT	240	

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dow Agrosciences LLC Patent Department
;; STREET: 9330 Zionsville Road
;; CITY: Indianapolis
;; STATE: Indiana
;; COUNTRY: USA
;; ZIP: 46268
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC Compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/036,987A
;; FILING DATE: 09-MAR-1998
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stuart, Donald R
;; REGISTRATION NUMBER: 28,479
;; REFERENCE/DOCKET NUMBER: 50,608
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317)337-4816
;; TELEFAX: (317)337-4847
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4928 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-036-987A-5

Query Match 6.1%; Score 83; DB 4; Length 4928;
Best Local Similarity 23.7%; Pred. No. 39;
Matches 62; Conservative 34; Mismatches 96; Indels 70; Gaps 10;

QY 29 QWRIRTESEVTA-DELAITIDGLIGEDSERLTGTAALSTVPSVLHEVRIMLDQYWPSPVPH 87
Db 1228 ENHRKALLGTAGDDLAIVGDPSPESVRATAREFA-----TLDEFRAAVSDSDVPAPGS 1281

QY 88 VL-----IIPGVRTGIPLLVDPNPKVGD--RIVNCLAAAYDRFRKAAIIVVDFGSSIC 137
Db 1282 VLVAAMSAEEVEGG-----SLPSRAQESTSDLLALVQSWLADERFAESQLVVVTRAAYS 1335

QY 138 VD-----VVSAGFELGGAIAAGVQVSSDAAAARSALRRVE--LARPR 179
Db 1336 ADSDSDVADLVGASSWGLLSSAQSENPGRFVLVDGTPESWQALPAAVRAGEPOLALRR 1395

QY 180 SVVGKNTVECMQAGAVFGFAGLVLDGLVGR-----IREDVSGFSVDHDAIVATGHTAPL- 233
Db 1396 GV-----ALVPLRLTLVREGGSPQLDGTGTLTGTTGALG 1433

QY 234 -LLPELHTVDHYDQHLTLQGLR 254
Db 1434 GVVARHLVEEHGIRRLVLAGRR 1455

RESULT 11
US-09-370-700-5
; Sequence 5, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Jan R
; APPLICANT: Waldron, Clive
; APPLICANT: Turner, Jan R
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987

;; EARLIER FILING DATE: 1998-03-09
;; NUMBER OF SEQ ID NOS: 39
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 5
;; LENGTH: 4928
;; TYPE: PRT
;; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-5

Query Match 6.1%; Score 83; DB 4; Length 4928;
Best Local Similarity 23.7%; Pred. No. 39;
Matches 62; Conservative 34; Mismatches 96; Indels 70; Gaps 10;

QY 29 QWRIRTESEVTA-DELAITIDGLIGEDSERLTGTAALSTVPSVLHEVRIMLDQYWPSPVPH 87
Db 1228 ENHRKALLGTAGDDLAIVGDPSPESVRATAREFA-----TLDEFRAAVSDSDVPAPGS 1281

QY 88 VL-----IIPGVRTGIPLLVDPNPKVGD--RIVNCLAAAYDRFRKAAIIVVDFGSSIC 137
Db 1282 VLVAAMSAEEVEGG-----SLPSRAQESTSDLLALVQSWLADERFAESQLVVVTRAAYS 1335

QY 138 VD-----VVSAGFELGGAIAAGVQVSSDAAAARSALRRVE--LARPR 179
Db 1336 ADSDSDVADLVGASSWGLLSSAQSENPGRFVLVDGTPESWQALPAAVRAGEPOLALRR 1395

QY 180 SVVGKNTVECMQAGAVFGFAGLVLDGLVGR-----IREDVSGFSVDHDAIVATGHTAPL- 233
Db 1396 GV-----ALVPLRLTLVREGGSPQLDGTGTLTGTTGALG 1433

QY 234 -LLPELHTVDHYDQHLTLQGLR 254
Db 1434 GVVARHLVEEHGIRRLVLAGRR 1455

RESULT 12
US-09-335-409-5
; Sequence 5, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-5

Query Match 6.1%; Score 83; DB 3; Length 7257;
Best Local Similarity 21.5%; Pred. No. 72;
Matches 47; Conservative 29; Mismatches 87; Indels 56; Gaps 8;

QY 57 RLGTAAALSTVPSVLHEVRIMLDQYWPSPVPHVLEIEPGVGTGIPLLVDPNPKVGDRIVC 116
Db 3612 RLDTQYAOQALFALEYALALWRSWGPVPHVLIHSHI-----GELVAC 3656

QY 117 LAAYDRFRKAAIIVVDFGSSICVVDVSAKGEFLGGAIAFGVQV---SSDAAAARSAA--LR 171
Db 3657 VAGV-----FSLDQAVRLVAARGRLMQALPAGAMVAIAASEAEEVAASVAPHAA 3705

QY 172 RVELARPRS-----VVGKNTVECMQAGAVFGAG-----LYDGLV---GRI 209
Db 3706 TVSTAAVNGPDVAVIAGAEVQVLAALGATFAARGIRTKRLAVSHAFSPMLDPMLEDFORV 3765


```

; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in
; FILE REFERENCE: Gram-Positive Microorganisms
; CURRENT APPLICATION NUMBER: US/09/462,844
; CURRENT FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14786
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305344.0
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Escherichia coli
; ORGANISM: Escherichia coli
; US-09-462-844-4

Query Match          6.1%; Score 83.5; DB 4; Length 323;
Best Local Similarity 25.6%; Pred. No. 0.43;
Matches 41; Conservative 27; Mismatches 51; Indels 41; Gaps 7;

QY 96 TGIPLVDPNPKVGDRIVNCIAAYDRFRKAA-----IVVDFGSS--ICVDVVSAAKGEFLG 149
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 56 TVIEITLKPAAIDVMR-----DALQKAGPEPMLQNFSGSHDVMRMPPAEGETGG 107

QY 150 GATAPGVQSSDAAARSALRVELARPSVVKNTVECMQAGVGFAGLVLD----- 203
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 108 QVLGSQVLKVINESTNONAAVKRIEFGPS--VG---ADLAQTGAMALMAALLSILVYVG 162

QY 204 -----GLVGRIREDV-----SGFSVDHDAIVAT 227
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 163 FFEWRLAGVVALAHADVITLGLSLFHLIEDIIVAS 202

RESULT 8
US-09-318-448-21
; Sequence 21, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-318-448-21

Query Match          6.1%; Score 83.5; DB 4; Length 344;
Best Local Similarity 21.9%; Pred. No. 0.48;
Matches 46; Conservative 30; Mismatches 65; Indels 69; Gaps 8;

QY 13 VGLLS-GMKEHAKVQVQWRIRTESEVTDELATIDGLGE-----DSERLTGTAAL 64
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 88 VVGINSETIMKPAISIEEELNLKLNDD--NVDGLLVQLPLPEHIDERRICNAVSP 144

QY 65 STVPSVLHEV---RIMLDQY--WPSVPHVLIIEGVTGIPLVND----- 104
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 145 DKVDGDFHVINVGRCMLDQYSMLPATPWGWEIKRTGPTLGNVWVAGRSKNVGMPIA 204

QY 105 -----PKE-----VGADRIVNCIAAYDRFRKAA 127
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 205 MLLHTDGAHERPGDGTVTISHRYPKEQLKKHTILADIVISAAGIPNLITA-DMIKEGA 263
```

```

QY 128 IVVDFGSSICVDVVSAAKGEFLGAIAPGVQ 157
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 264 AVIDVGINRVHDPVTAKPKLVGDVDFEGVR 293

RESULT 9
US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-320-878-1

Query Match          6.1%; Score 83.5; DB 3; Length 4551;
Best Local Similarity 27.4%; Pred. No. 30;
Matches 46; Conservative 28; Mismatches 65; Indels 29; Gaps 11;

QY 83 PSYPHVLIIEPGVTGIPLVDPNPKVGDRIVNCIA-----AYDRFRKAAIIVDFGSS 135
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 406 PAIP--FEELNLRVNTYEILPWEPEHGDQRMVGVSSFGMGGTNAHVVLEAPGVVE-GAS 462

QY 136 ICVDVVSAAKGEFLGAIAPGVQVSSDAAARSALRVELARPSVVKNTVECMQAGV 195
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 463 VVESTVG--GSAYGGVVPVW-VSAKSAALDAQIERL-----AASFASRDRTDGVDAV 514

QY 196 FGFAGLVLD-GLVGRIREDVSG--FSVDHDAIVATGH---TAPLLLE 237
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 515 --DAGAVDAGAVARV---LAGGRAQFHRVAVVVGSGPDDLAAALAAPE 557

RESULT 10
US-09-036-987A-5
; Sequence 5, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF INVENTIONS: Production
; NUMBER OF SEQUENCES: 39
```

REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3724 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-227C-10

Query Match 6.4%; Score 87; DB 2; Length 3724;
Best Local Similarity 24.8%; Pred. No. 8.7;
Matches 53; Conservative 27; Mismatches 52; Indels 82; Gaps 12;

QY 75 RIMLDQYWPSPHVLIEP-----GVRTG-----IPLLVDNPKVGAADRVNCLAA 120
Db 1637 RLILETSWESPERAGIDPVELRGSTGVFGTNGQHYVPLLODGDEN-----F 1684
QY 121 DRFRKAAIVDFGSSICVDVVSAGKEFLGGAIPGVQVSSDAAAARSA-----ALRR-- 172
Db 1685 DGY-----IATGNS--ASVMSGRLSYVFGLEGPATVTDTCASLAALHLAVQSLRRGE 1736
QY 173 -----VELARPSVV--GKNTVECMQAGAVFGAGLVGD-----LVGR 208
Db 1737 CDYALAGGATVMSTPEMLVEFARQAVSPDGRSKAFARADGV-----GLAEGAGMLLYER 1792
QY 209 IRED-----VSGFSVDHDAIVAATGTAP 232
Db 1793 LSEAKKGHPVLAVVRSVAVNODGA--SNGLTAP 1824

RESULT 5
US-08-804-198-4
Sequence 4, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3724 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide

US-08-804-198-4

Query Match 6.4%; Score 87; DB 2; Length 3724;
Best Local Similarity 24.8%; Pred. No. 8.7;
Matches 53; Conservative 27; Mismatches 52; Indels 82; Gaps 12;

QY 75 RIMLDQYWPSPHVLIEP-----GVRTG-----IPLLVDNPKVGAADRVNCLAA 120
Db 1637 RLILETSWESPERAGIDPVELRGSTGVFGTNGQHYVPLLODGDEN-----F 1684
QY 121 DRFRKAAIVDFGSSICVDVVSAGKEFLGGAIPGVQVSSDAAAARSA-----ALRR-- 172
Db 1685 DGY-----IATGNS--ASVMSGRLSYVFGLEGPATVTDTCASLAALHLAVQSLRRGE 1736
QY 173 -----VELARPSVV--GKNTVECMQAGAVFGAGLVGD-----LVGR 208
Db 1737 CDYALAGGATVMSTPEMLVEFARQAVSPDGRSKAFARADGV-----GLAEGAGMLLYER 1792
QY 209 IRED-----VSGFSVDHDAIVAATGTAP 232
Db 1793 LSEAKKGHPVLAVVRSVAVNODGA--SNGLTAP 1824

RESULT 6
US-09-327-681-6
Sequence 6, Application US/09327681
Patent No. 6284495
GENERAL INFORMATION:
APPLICANT: Sato, Katsuaki
APPLICANT: Usuda, Yoshihiro
TITLE OF INVENTION: Method for producing nucleic acid substances
FILE REFERENCE: OP853
CURRENT APPLICATION NUMBER: US/09/327,681
CURRENT FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: JP 10-165704
PRIOR FILING DATE: 1998-06-12
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 6
LENGTH: 285
TYPE: PRT
ORGANISM: Bacillus subtilis
US-09-327-681-6

Query Match 6.1%; Score 84; DB 4; Length 285;
Best Local Similarity 22.6%; Pred. No. 0.31;
Matches 60; Conservative 37; Mismatches 100; Indels 68; Gaps 12;

QY 12 TVVGLSGMKEHAKVWQWRIRTESEVTADELALTIDGLIGDSERLGTAAALST---V 67
Db 61 TVPGNAGGVKYPKMQ-----AEAEFVQTLGOSLANPERILPGYVYLLDLGK 111
QY 68 PSVLHEVRIMLDQYWPSPHVLIEPVRTGIPLLVDNPKVGAADRVNCLAAAYDRFKAA 127
Db 112 PSVLSKVGKLEASVFAEREIDVMTVATKGIPL-----AY-----AA 148
QY 128 IVDGSSICVDVVSAGKEFLGGAIPGVQVSSDAAAARSAALRRVELARPSVVGKTV 187
Db 149 -----ASYLNPVVIVRKD---NKVTEGSTSVINTVSGSSNRIOPTMSLAKRSMKTGSNVL 200
QY 188 ---ECMOAGAVFGAGLVGDGLVGRIRE---DVSGFSVDHDAIVAATGTAPLLLPETHV 241
Db 201 IIDDFMKAG-----GTINGMINLLDEFNANVAGIG----VLVEAGVDERL-----V 243
QY 242 DHDQHLTLGLRLVFERNLEVRQ 266
Db 244 DEYMSLLTSLTNMK-EKSIEIQNG 267

RESULT 7
US-09-462-844-4
Sequence 4, Application US/09462844
Patent No. 6258563

```

; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,761A
; FILING DATE: 28-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: UALB-03356
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8336
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-09-085-761A-34

Query Match      6.98; Score 95; DB 4; Length 140;
Best Local Similarity 26.6%; Pred. No. 0.0055;
Matches 47; Conservative 17; Mismatches 37; Indels 76; Gaps 10;

QY   78 LDQYWP-----SVPHVL-----IEPGVRTGIPLVDNPKPEGADRIYNCLAAAYDRFR 124
     :| |||| :||| :||| |::|::| :||| |
Db   22 IDMYWGRLDGCAEPHQREAFWHIDLAKRTGKPLMHNQ-----ADRDV----- 68

QY   125 KAAIVDFGSSICVVVSAGKEFLGGIAIPGVQV-----SSDAAAARSAAALRRVELAPRS 180
     ::| :| | :| | :| | :| | :| | :| |
Db   69 -----LDVLRAEG-----APDTVLHCFSSDAAMART----- 95

QY   181 VVGKNITVECMQAGAVFGAGLVLDGLVCR-IREDVSGFSVDHDVAIVATGHTAPILLP 236
     :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db   96 -----CVDAGWLISLGSVTSFRTARELREAVPLMPVEQ--LLVET--DAPYLTP 140

RESULT 4
US-08-804-227C-10
; Sequence 10, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kubstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas. G.
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.94596 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-5
Perfect score: 1367
Sequence: 1 MLLAIDVRNTHVTVVLLSGM.....LRLVFERNLEVQGRLEKTAR 272

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Issued Patents AA:*
1: /cgn2.6/ptodata/1/iaa/5A-COMB.pep.*
2: /cgn2.6/ptodata/1/iaa/5B-COMB.pep.*
3: /cgn2.6/ptodata/1/iaa/6A-COMB.pep.*
4: /cgn2.6/ptodata/1/iaa/6B-COMB.pep.*
5: /cgn2.6/ptodata/1/iaa/PCTUS-COMB.pep.*
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	8.4	338	4	US-09-199-637A-47
2	95	6.9	140	3	US-09-053-197A-34
3	95	6.9	140	4	US-09-085-761A-34
4	87	6.4	3724	2	US-08-804-227C-10
5	87	6.4	3724	2	US-08-804-198-4
6	84	6.1	285	4	US-09-327-681-6
7	83.5	6.1	323	4	US-09-462-844-4
8	83.5	6.1	344	4	US-09-318-448-21
9	83.5	6.1	4551	3	US-09-320-878-1
10	83	6.1	4928	4	US-09-036-987A-5
11	83	6.1	4928	4	US-09-370-700-5
12	83	6.1	7257	3	US-09-335-409-5
13	83	6.1	7257	4	US-09-568-102-5
14	83	6.1	7257	4	US-09-567-969-5
15	83	6.1	7257	4	US-09-568-480-5
16	83	6.1	7257	4	US-09-568-486-5
17	83	6.1	7257	4	US-09-568-472-5
18	83	6.1	7257	4	US-09-567-899-5
19	81.5	6.0	648	4	US-09-134-001C-5161
20	81	5.9	1580	2	US-08-804-227C-11
21	81	5.9	1580	2	US-08-804-198-5
22	80.5	5.9	3170	2	US-07-642-734C-5
23	80.5	5.9	3170	3	US-08-439-009A-5
24	79.5	5.8	4613	4	US-09-105-537-31
25	79.5	5.8	11877	4	US-09-105-537-6
26	78.5	5.7	654	4	US-09-620-412C-341
27	78.5	5.7	1752	4	US-09-556-877-180

```

28 78.5 5.7 1752 4 US-09-620-412C-180 Sequence 180, App
29 78 317 4 US-09-199-637A-433 Sequence 433, App
30 78 572 2 US-08-453-848-9 Sequence 9, Appl
31 78 572 4 US-09-169-027-9 Sequence 9, Appl
32 78 636 2 US-08-871-266B-2 Sequence 2, Appl
33 78 572 636 2 US-08-819-458A-2 Sequence 2, Appl
34 78 636 2 US-09-018-864A-2 Sequence 2, Appl
35 78 636 3 US-08-871-267B-2 Sequence 2, Appl
36 78 636 4 US-09-618-419-2 Sequence 2, Appl
37 77.5 5.7 712 4 US-09-708-426-9 Sequence 9, Appl
38 77.5 5.7 843 4 US-09-361-631-2 Sequence 2, Appl
39 77 3729 2 US-08-804-227C-4 Sequence 4, Appl
40 76.5 5.6 733 4 US-09-192-983-6 Sequence 6, Appl
41 76 417 1 US-08-464-523B-23 Sequence 23, Appl
42 75.5 5.5 322 2 US-08-932-978-2 Sequence 2, Appl
43 75.5 5.5 423 2 US-08-846-762-74 Sequence 74, Appl
44 75.5 5.5 489 4 US-08-942-012B-31 Sequence 31, Appl
45 75.5 5.5 3519 4 US-09-428-517-4 Sequence 4, Appl

```

ALIGNMENTS

```

RESULT 1
US-09-199-637A-47
; Sequence 47, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199.637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066.517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-47

```

```

Query Match      8.4%; Score 115; DB 4; Length 338;
Best Local Similarity 27.3%; Pred. No. 0.00012;
Matches 63; Conservative 23; Mismatches 97; Indels 48; Gaps 10;

Qy 47 IDGLIGEDSERLT-----GTAISTVPSVLHEVRIMLDQYWP-----SVPHV- 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 IAGALGDDGCRLLVYVGLVQEAARIEYPPHYVAGEDSTHLAQPDQRFVHLLGNSMPAN 157
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 89 -----LIEPGVRTGILLVDNPKVEGADRVNCLAAAYDRFRKAAIVVDFGSSICV 138
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 GVQCAEKVRHQDGGARANVPRGAGEPAERGATR----MADHIFLEAADAV-LGLVCG 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 139 DWSAKGE-----FLGGAIAPGVQVSSDAARAASAA--LRRVELARPSRVVGKNT-V 187
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 213 RVITAGLGEIRCTQRRYLPGVAPGIRVAGDDCVRHVVADLDRHLHFAAMRAAEQPVTD 272
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 188 ECMQAGAVFGAGLVGDLV---GRIRE-DVSGFSVDHDAIVATGHTAPLL 234
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 273 DDLVFEALRGKGDDGDSAVDRGREREAREAGGRRRCQAAFEVAGHQDRL 323
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 2

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 4.40539 seconds
(without alignments)
2061.866 Million cell updates/sec

Title: US-09-813-453a-57

Perfect score: 1093

Sequence: 1 MSEKLVAVDICNTSVNIGIF.....IGLVEELVRRITQELGQKTR 219

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	518	47.4	233	1 YACB_BACSU	P37564 bacillus su
2	130.5	11.9	267	1 BAF_BORPE	Q45338 bordetella
3	95.5	8.7	402	1 OPS4_CANAL	P46596 candida alb
4	94	8.6	253	1 HIS6_AQUAE	O66567 aquifex ae
5	91	8.3	427	1 Y428_METJA	Q57871 methanococc
6	88	8.1	394	1 THIL_CHRVI	P45369 chromatiu
7	88	8.1	404	1 AK1_BACSU	Q04795 bacillus su
8	87.5	8.0	328	1 GLK_STAXY	Q56198 staphylococ
9	87.5	8.0	484	1 Y551_METJA	Q58846 methanococc
10	87	8.0	428	1 PURA_CLOAB	Q97d87 clostridium
11	86	7.9	836	1 V626_BPM15	Q05233 mycobacteri
12	85.5	7.8	497	1 DLHD_MANSE	O18480 manduca sex
13	84.5	7.7	356	1 BUK_CLOPE	Q92ne5 clostridium
14	84	7.7	393	1 FLAA_PSEAE	P21184 pseudomonas
15	84	7.7	393	1 THIL_ALCEU	P14611 alcaligenes
16	84	7.7	420	1 FTSA_ECOLI	P06137 escherichia
17	84	7.7	603	1 LPPA_SYNY3	P74751 synchocyst
18	84	7.7	654	1 DNAK_CHLMU	P56836 chlamydia m
19	84	7.7	675	1 CMC2_HUMAN	Q9ujs0 homo sapien
20	83.5	7.6	428	1 PURA_CLOPE	Q8xh63 clostridium
21	83.5	7.6	740	1 PAB1_ANASP	P58565 anabaena sp
22	83.5	7.6	740	1 PSAB_ANAVA	P31088 anabaena va
23	83	7.6	312	1 MDH_ECOLI	P06994 escherichia
24	82	7.5	394	1 THIL_THIVI	P45363 thlocystis
25	82	7.5	604	1 DNAB_BACME	P05646 bacillus me
26	82	7.5	607	1 DNAB_BACST	Q45551 bacillus st
27	82	7.5	695	1 CMC1_DROME	Q9va73 drosophila
28	81.5	7.5	311	1 MRAN_CAUCR	O9rqj6 caulobacter
29	81.5	7.5	501	1 DLDH_PEA	P31023 pisum sativ
30	81.5	7.5	613	1 DNAB_BACHD	Q9kd72 bacillus ha
31	81	7.4	492	1 IMDH_STRPY	P50099 streptococc
32	81	7.4	610	1 DNAB_STAAM	Q99tr7 staphylococ
33	81	7.4	610	1 DNAB_STAAM	P45554 staphylococ

RESULT 1

YACB_BACSU
ID YACB_BACSU STANDARD; PRT: 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RT Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RT Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RT Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
RT Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RT Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RT Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RT Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RT Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RT Ghm S.Y., Glaser P., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RT Guiseppi G., Guy B.J., Haga K., Haiech J., Gollightly E.J., Grandi G.,
RT Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RT Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RT Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RT Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RT Lee S.M., Levine A., Liu H., Masuda S., Mauel J., Medigue C.,
RT Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RT Noone D., O'Reilly T.M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RT Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RT Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RT Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RT Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RT Seguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RT Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RT Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RT Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RT Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RT Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RT Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

ALIGNMENTS

34	81	7.4	659	1	DNAB_CHLTR	P17821 chlamydia t
35	80.5	7.4	219	1	Y06K_BPT4	P13318 bacterioph
36	80.5	7.4	333	1	DRPB_BACFI	P94311 bacillus fi
37	80.5	7.4	730	1	PSAB_SYNY3	P29255 synchocyst
38	80	7.3	252	1	HIS6_RHOSH	P50937 rhodobacter
39	80	7.3	393	1	CYSK_YEAST	P53206 saccharomyc
40	80	7.3	434	1	HISX_SYNY3	P73058 synchocyst
41	80	7.3	610	1	DNAB_BACSU	P17820 bacillus su
42	80	7.3	615	1	DNAB_THETH	Q56235 thermus the
43	80	7.3	676	1	CMC2_MOUSE	Q9qxx4 mus musculu
44	79.5	7.3	336	1	Y508_ARCFU	O29742 archaeoglob
45	79.5	7.3	534	1	VL2_HPV37	O80905 human papil

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D26185; BAA05305.1; -
 CC EMBL; Z99104; CAB11846.1; -
 CC Subtilist; BG10133; yacB.
 CC InterPro; IPR004619; Baf.
 CC Pfam; PF03309; Bvg_acc_factor; 1.
 CC TIGRFAMS; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

 Query Match 47.4%; Score 518; DB 1; Length 233;
 Best Local Similarity 45.6%; Pred. No. 5.2e-34;
 Matches 98; Conservative 45; Mismatches 72; Indels 0; Gaps 0;

 Db 5 LVAVDIGNTVNIGIFEGEKLLANWHLGSAQORMADEYASLLGLLQHAGIHPEELNRVI 64
 2 LVIDVGNNTVLGVYHDGKLEWHRIETSRHKTEDEFGMLRSLFDHSLMFQIDGII 61
 QY 65 MCSVVPPLTTTFEEVFKSYFKAAPLVVVGAGIKSGVKRMDNPREVGADRIVNAARVLY 124
 Db 62 ISSVVPPIFALEBCTKTFHIEPIQVPGMKTGLNIKVDNPKVGDRIVNAVAIHLY 121
 QY 125 PGACIIIVDMGTATFTLSEGGAYIGGAITAPGATSAQAIAEKTSKLPKIEIIRPAKVIG 184
 Db 122 GNPLIVVDGATATYCIDENKQYMGGAITAPGITISTEALYSRAAKLPRIETRPDNIIG 181
 QY 185 SNTVSAMQSGIIVGYGLVEELVRRITQELGQKTR 219
 Db 182 KNTVSAMQSGIIVGYGVQVEGIVKRMKQAKQDPR 216

 RESULT 2
 BAF_BORPE STANDARD; PRT; 267 AA.
 AC Q45338; Q45373; -
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Bvg accessory factor.
 GN BAF.
 OS Bordetella pertussis.
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 Bordetella.
 NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP504;
 RX MEDLINE=95325323; PubMed=7601846;
 RA Deshazer D., Wood G.E., Friedman R.L.;
 RT "Identification of a Bordetella pertussis regulatory factor required
 RT for transcription of the pertussis toxin operon in Escherichia
 RT coli";
 RL J. Bacteriol. 177:3801-3807(1995).
 RN [2]
 RP SEQUENCE OF 1-38 FROM N.A.
 RC STRAIN=BP504;
 RA Wood G.E., Friedman R.L.;
 RT "Identification of a bira homolog in Bordetella pertussis";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 239-267 FROM N.A.
 RC STRAIN=BP536;
 RX MEDLINE=96419162; PubMed=8821935;
 RA Allen A.G., Maskell D.J.;
 RT "The identification, cloning and mutagenesis of a genetic locus

required for lipopolysaccharide biosynthesis in Bordetella
 pertussis";
 Mol. Microbiol. 19:37-52(1996).
 CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
 CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
 CC RNA POLYMERASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U12020; AAA75361.1; -
 CC EMBL; AF016461; AAC68834.1; -
 CC EMBL; X90711; CAA62242.1; -
 CC InterPro; IPR004619; Baf.
 CC Pfam; PF03309; Bvg_acc_factor; 1.
 KW Transcription regulation; Activator.
 SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

 Query Match 11.9%; Score 130.5; DB 1; Length 267;
 Best Local Similarity 23.6%; Pred. No. 0.0019;
 Matches 52; Conservative 36; Mismatches 87; Indels 45; Gaps 8;

 QY 5 LVAVDIGNTVNIGIFEGE-----KLLANWHLGSAQORMADEYASLL 46
 Db 2 IILIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRW-LATLPRRQALGVNV 60
 QY 47 LGLLQHAGIHPEELNRVIMCSVVPPLTTTFEEVFKSYFKAAPLVVVGAGIKSGVKRMDNP 106
 Db 61 AGLARGEAI--AATLRAGGCDI-----RWLRQAQPLAM--GLRNGYR----NP 99
 QY 107 REVGADR---IVNAAARVLYPGACIIIVDMGTATFTLSEGGAYTGGAIAPGATSAQA 163
 Db 100 DQLGADRWACMVGVLARQPSVHPPLILVASFGTATLDTIGPDNVFPGILPGPMARGA 159
 QY 164 IAEKTSKLPKIEIIRPAKVIGSNTVSAMQSGIIVGYGVIGLV 203
 Db 160 LAYGTAHLPLADGLVADYPI--DTHQAIASGIAAAQAAGAI 197

 RESULT 3
 OPS4_CANAL STANDARD; PRT; 402 AA.
 ID OPS4_CANAL
 AC P46596;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Opaque-phase-specific protein OP4 precursor.
 GN OPS4.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WO-1;
 RX MEDLINE=93239284; PubMed=8478072;
 RA Morrow B., Srikantha T., Anderson J., Soll D.R.;
 RT "Coordinate regulation of two opaque-phase-specific genes during
 RT white-opaque switching in Candida albicans";
 RL Infect. Immun. 61:1823-1828(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR PROSITE; PS00737; THIOLEASE_2; 1.
KW Transferase; Acyltransferase; PHB biosynthesis.
FT ACT_SITE 89 SUBSTRATE BINDING (BY SIMILARITY).
FT ACT_SITE 380 BASE (BY SIMILARITY).
SQ SEQUENCE 394 AA; 41053 MW; 95AC363A270FA14A CRC64;

Query Match 8.1%; Score 88; DB 1; Length 394;
Best Local Similarity 29.9%; Pred.No. 6.3;
Matches 32; Conservative 17; Mismatches 36; Indels 22; Gaps

QY 1 MSEKLVAIVDIGNTSVNGFFEGEKKLANWHLGVSVAORMADEVASLL-GLLQHAGIHPEE 59
DB 1 MSENVIVIDGRSA -IGTFG-----GSLSSLSATELGTVAGLRGLARTGLAPEQ 48

QY 60 LNRVMCSVV-----PPLTTTTFEFKSFKAAPL--VVGAGIKS 97
DB :: ||: |: |:: ||: ||: :||| :||| :||: ||:

49 IDEVLGVQLTAGVGNPARQTTLHAGLPSPAMTIINKVCSGSLKA 95
DB :: ||: |: |:: ||: ||: :||| :||| :||: ||:

RESULT 7
AKI_BACSU STANDARD; PRT: 404 AA.

AC Q04795; O31759;
DT 01-OCT-1993 (Rel. 27, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aspartokinase I (EC 2.7.2.4) (Aspartokinase I) (Aspartate kinase 1)
DE [contains: Aspartokinase I alpha subunit; Aspartokinase I beta subunit].
GN DAPG OR USSD.
OS Bacillus subtilis.
OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
[1]
SEQUENCE FROM N.A.
RP STRAIN=168;
RC MEDLINE=93252813; PubMed=8098035;
RX Chen N.-Y., Jiang S.-O., Klein D.A., Paulus H.;
RT "Organization and nucleotide sequence of the Bacillus subtilis diamino-pimelate operon, a cluster of genes encoding the first three enzymes of diaminopimelate synthesis and dipicolinate synthase.";
RL J. Biol. Chem. 268:9448-9465(1993).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Boletín A., Borcherdt S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Broullet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dunsterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Funas S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goiffau A., Golightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holtsappel S., Hosono S., Hulio M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Ouéga B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpsater P., Tognoni A., Tosato V., Uchiyama S., Vandebol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wibpat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 subsp. n.":
 RL Nature 390:249-256(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
 aspartate.
 CC -!- ENZYME REGULATION: DIAMINOPIMELATE-SENSITIVE.
 CC -!- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
 FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
 CC -!- SUBUNIT: TETRAMER CONSISTING OF TWO ALPHA (CATALYTIC) AND TWO BETA
 (FUNCTION NOT KNOWN) CHAINS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L08471; AA22384.1; -
 DR EMBL: Z91112; CAB13549.1; -
 DR PIR: S34600; S34600.
 DR PIR: C46665; C46665.
 DR Subtilist; BG10784; dapG.
 DR InterPro: IPR001048; Aa_kinase.
 DR InterPro: IPR005260; Asp_kin_monofn.
 DR InterPro: IPR001341; Aspartate_kinase.
 DR InterPro: IPR001057; Glu_5kinase.
 DR Pfam: PF00696; aakinase; 1.
 DR PRINTS: PR00474; GLUSKINASE.
 DR TIGRFAMs: TIGR00656; asp_kin_monofn; 1.
 DR TIGRFAMs: TIGR00857; asp_kinases; 1.
 DR PROSITE: PS00324; ASPARTOKINASE; 1.
 KW Transferase; Kinase; Diaminopimelate biosynthesis;
 Lysine biosynthesis; Complete proteome.
 FT CHAIN 1 404 ASPARTOKINASE I ALPHA SUBUNIT.
 FT CHAIN 245 404 ASPARTOKINASE I BETA SUBUNIT
 FT CONFLICT 399 399 E -> V (IN REF. 1).
 FT SEQUENCE 404 AA; 42978 MW; 49A6DA70D70047C4 CRC64;
 Y 4 KLVAVDIGNTSVN-----IG-----IFEGEKLNNHLSGVAQRMADAYAS-LLGL 49
 Db 2 KIIVQFGGTSVKDKGRKRLALGHKEAISEGYKVVV---VVSAMGRKGDPTVATDLSLGL 58
 QY 50 L--QHAGTHPELNRMVMSVPPPLTTTFEEVFKSYF-----KAAPLVVAGIKSGVKVR 102
 Db 59 LYGDQSASPREQDILLSCG-----ETISSVVFTSMLLDNGVKAAL---TGAQAGFLTN 110
 QY 103 MDNPR-----EVCADRVNAAAR---VLYPGACIIIVDMGTATFTDLSLSEGG-----AYLGG 151
 Db 111 DQHTNAKIEMKPERLFSVLANHDVAVVAGFCGATEKGGDTT---TIGRGGSDTSAAALGA 167
 QY 152 AI-----APGIATSAQAIAEKTSLPKI-----EI-----IRPAKVGISNTVS-AMOS 193
 Db 168 AVDAEYIDIFTVEGVMTPADPRVVENAKPLPVVTVTEICNLAYQCAKVISPRAVEIAMQA 227
 QY 194 GI 195
 Db 228 KV 229
 RESULT 8
 GLK_STAXY STANDARD; PRT; 328 AA.
 AC Q56198;
 DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glucokinase (EC 2.7.1.2) (Glucose kinase).
 GN GLKA.
 OS *Staphylococcus xylosum*.
 OC Bacteria; Firmicutes; Bacillales; *Staphylococcus*.
 OX NCBI_TaxID=1288;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 20267 / Isolate C2A;
 RX MEDLINE=96042090; PubMed=7592379;
 RA Wagner E., Marcandier S., Egeter O., Deutscher J., Goetz F.,
 Brueckner R.;
 RA "Glucose kinase-dependent catabolite repression in *Staphylococcus*
xylosum."; J. Bacteriol. 177:6144-6152(1995).
 RL J. Bacteriol. 177:6144-6152(1995).
 CC -!- CATALYTIC ACTIVITY: ATP + D-glucose = ADP + D-glucose 6-phosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLK) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X84332; CA59069.1; -
 DR InterPro: IPR000600; ROK_family.
 DR InterPro: IPR004654; ROK_glca_fam.
 DR Pfam: PF00480; ROK; 1.
 DR TIGRFAMs: TIGR00744; ROK_glca_fam; 1.
 DR PROSITE: PS01125; ROK; 1.
 DR TRANSFERASE; Kinase; Glycolysis; ATP-binding.
 KW Transferase; Kinase; Glycolysis; ATP-binding.
 SQ SEQUENCE 328 AA; 35032 MW; 9BF622D66AF1B5DF CRC64;
 Query Match 8.0%; Score 87.5; DB 1; Length 328;
 Best Local Similarity 22.8%; Pred. No. 5.6;
 Matches 56; Conservative 25; Mismatches 106; Indels 59; Gaps 8;
 QY 1 MSEKLAVDIGNTSVNIIGIF-EGEKLLNNHLSGVAQRMADAYASLLL-----GLLQHAG 54
 Db 1 MNKIILAADIGTGTCKLIGFDENLNLSKW---SIDTSDTGTGYYLLKNNIYDSFIQHD 57
 QY 55 IIPPEELNRVIMSVVPPPLTTTFEE-----VFKSYFKAAPLVVAGIKSGVKVRMDNPREV 109
 Db 58 KSDNTFSDVILGIGIGYVPGVNFETGEVNGAVNLVWKG-----VNYRDIKQFVDCP--V 110
 QY 110 GADRVNAAARVLYPGACIIIVDMGTATFTDLSLSEGGAYIGGAIAPG-----156
 Db 111 YVDNANVAALGEKKHKGAGGADVVVAITLGTGLGGGIISNGEIVHNGSGAEIGHFRV 170
 QY 157 -----IATSAQAIA-----EXTSKLPKTEIIRPAKVGISNTVSAM 191
 Db 171 DHDQFKCKNGKSGCIETVASATGVVNLVNFYFKLTFKSSILQLKDNKVSAAKAVFDA 230
 QY 192 QSGIYF 197
 Db 231 KAGDQF 236
 RESULT 9
 YESL_METJA STANDARD; PRT; 484 AA.
 ID YESL_METJA
 AC Q58846;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ1451.
 GN MJ1451.
 OS *Methanococcus jannaschii*.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

Search completed: June 24, 2003, 22:11:54
Job time : 5.40539 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 19.2642 seconds
(without alignments)
2342.388 Million cell updates/sec

Title: US-09-813-453A-57
Perfect score: 1093
Sequence: 1 MSEKLVAVDIGNTSVNIIGIF.....IGLVEELVRRIQTELQKTR 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	536	49.0	255	Q8r7m2	thermoanaer
2	533.5	48.8	273	Q97eb4	clostridium
3	527	48.2	254	Q9kgh5	bacillus ha
4	508	46.5	258	2 Q9f985	Q9f985 bacillus st
5	506.5	46.3	259	Q8xhl5	Q8xhl5 clostridium
6	492.5	45.1	265	16 Q9X8N6	Q9x8n6 streptomyce
7	489	44.7	259	16 Q8YAC5	Q8yac5 listeria mo
8	483	44.2	259	16 Q92F54	Q92f54 listeria in
9	478	43.7	261	16 Q9A6Z1	Q9a6z1 caulobacteri
10	373	34.1	274	16 Q9CD56	Q9cd56 mycobacteri
11	371	33.9	272	16 Q06282	Q06282 mycobacteri
12	348.5	31.9	212	2 Q32514	Q32514 desulfovibr
13	333.5	30.5	262	16 Q9RX54	Q9rx54 deinococcus
14	330	30.2	256	16 Q8RFE4	Q8rfe4 fusobacteri
15	310	28.4	273	16 Q83446	Q83446 treponema p
16	279.5	25.6	246	16 Q9WZY5	Q9wzy5 thermotoga

17	190	17.4	262	16	O51477	borrelia bu
18	169	15.5	295	16	Q8Y2M4	Q8y2m4 ralstonia s
19	160	14.6	276	16	Q8YQD7	Q8yqd7 anabaena sp
20	152	13.9	592	16	Q9JW17	Q9jw17 neisseria m
21	150	13.7	592	16	Q9JXF1	Q9jxf1 neisseria m
22	144	13.2	257	16	P74045	P74045 synechocyst
23	143.5	13.1	242	16	Q9PC14	Q9pci14 xylella fas
24	136	12.4	248	16	Q9HWC1	Q9hwc1 pseudomonas
25	130	11.9	229	16	O67753	O67753 aquifex aeo
26	118	10.8	223	16	Q92KY6	Q92ky6 helicobacte
27	108	9.9	223	16	O25533	O25533 helicobacte
28	96	8.8	322	2	Q8VVB8	Q8vvb8 streptococc
29	94.5	8.6	467	17	Q8ZUT2	Q8zut2 pyrobaculum
30	94	8.6	393	16	Q8Y2Q9	Q8y2q9 ralstonia s
31	94	8.6	440	2	Q9FCQ4	Q9fcq4 pseudomonas
32	92	8.4	340	2	P72123	P72123 pseudomonas
33	91	8.3	56	2	P94305	P94305 bacillus ps
34	91	8.3	453	16	Q9KQ51	Q9kq51 vibrio chol
35	90.5	8.3	1047	16	Q8XRD0	Q8xrd0 ralstonia s
36	90	8.2	278	12	Q65032	Q65032 avian pneum
37	89.5	8.2	396	16	Q8YBR9	Q8ybr9 brucella me
38	89.5	8.2	420	16	Q8KPH0	Q8kph0 vibrio chol
39	89	8.1	420	16	Q8YJ30	Q8yj30 brucella me
40	88.5	8.1	282	2	Q9KH82	Q9kh82 escherichia
41	88.5	8.1	559	16	Q9RW79	Q9rw79 deinococcus
42	88.5	8.1	560	16	Q8YVS8	Q8yvs8 anabaena sp
43	88	8.1	340	2	O33928	O33928 pseudomonas
44	88	8.1	399	17	Q8ZX10	Q8zx10 pyrobaculum
45	88	8.1	509	4	Q9U107	Q9u107 homo sapien

ALIGNMENTS

RESULT 1

Q8R7M2	PRELIMINARY;	PRT;	255 AA.
ID	Q8R7M2		
AC	Q8R7M2;		
DT	01-JUN-2002 (TrEMBLrel. 21, Created)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Putative transcriptional regulator, homologs of Bvg accessory factor.		
DE	factor.		
GN	TTE2381.		
OS	Thermoanaerobacter tengcongensis.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;		
OC	Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.		
OX	NCBI_TaxID=119072;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MB4T / JCM11007;		
RX	MEDLINE=21992816; PubMed=11997336;		
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,		
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,		
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;		
RT	"A complete sequence of T. tengcongensis genome.";		
RL	Genome Res. 12:689-700(2002).		
DR	EMBL; AF013180; AAM25520.1;		
KW	Complete proteome; Hypothetical protein.		
SQ	SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;		

Query Match 49.0%; Score 536; DB 16; Length 255;
Best Local Similarity 48.3%; Pred. No. 1.6e-37;
Matches 101; Conservative 50; Mismatches 58; Indels 0; Gaps 0;

QY	5	LVAVDIGNTSVNIGIFEGEKLLANWHLGSAQVMADEYASLLGLLQHGAGHPPELNRYI 64
Db	2	LLAFDVGNTNIVMGVFKGKLLHSFRISFDKNTKYDEYGMVNLQGLYNGISLTEDDVI 61
QY	65	MCSSVPPLTTTTEEVKFSYKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNAARVLY 124
Db	62	ISSVVPPLMTLQVMSLKYFKTPVVGPGIKTKINIKYDNPKEVGADRIVNAAYELY 121

QY 125 PGACIIIVDMGTATFTDLSGGAYIGGAIAPGIATSAQAIAEKTSLPKRIEIRPAKVI 184
DB 122 GGPVIVIDGTATFTCAISEKGYLGIIAPGLMISADALFORTAKLPIDLTCKPPTVIN 181
QY 185 SNTVSAMQSGIYFGYIGLVEELVRRIQTE 213
DB 182 RNTVASMOSGIYGHVGMVDYIVTRMKGE 210

RESULT 2

Q97EB4 PRELIMINARY; PRT; 273 AA.
ID Q97EB4;
AC Q97EB4;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory factor.
DE factor.
GN CAC3200.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
NCBI_TaxID=1488;

SEQUENCE FROM N.A.

RP STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007815; AAK81136.1; -
DR InterPro: IPR004619; Baf.
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS: TIGR00671; baf; 1.
DR PROSITE: PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;

Query Match 48.8%; Score 533.5; DB 16; Length 273;

Best Local Similarity 47.7%; Pred. No. 2.9e-37;

Matches 103; Conservative 44; Mismatches 66; Indels 3; Gaps 1;

QY 5 LVAVDIGNTSVNIGIFEKGLLANHLSVQAORMADEYASLLGLLQHAGIHPEELNRVI 64
DB 13 ILVLDVGNNTVLGVYQDETLLVHWRLATSRKTEDEYAMTVRSLEDFHAGLQFQDIDGIV 72
QY 65 MCSVVPPLTTTTEEVFKSYFKAAPLVVAGIKSGVKVRMDNPREVGADRIVNAARVLY 124
DB 73 ISSVVPNMYSLEHMRKFKINPLVVGICITGINIKYDNPKEVGADRIVNAARVLY 132
QY 125 PGACIIIVDMGTATFTDLSGGAYIGGAIAPGIATSAQAIAEKTSLPKRIEIRPAKVI 184
DB 133 KRSLLIIDGTATFTCAVRENGDYLGAICPGIKYSSEALFEKAALPRVELIKPAYAIC 192

Query Match 48.8%; Score 533.5; DB 16; Length 273;

Best Local Similarity 47.7%; Pred. No. 2.9e-37;

Matches 103; Conservative 44; Mismatches 66; Indels 3; Gaps 1;

RESULT 3

Q9KGH5 PRELIMINARY; PRT; 254 AA.
ID Q9KGH5;
AC Q9KGH5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein BH0086.

Query Match 48.8%; Score 533.5; DB 16; Length 273;

Best Local Similarity 47.7%; Pred. No. 2.9e-37;

Matches 103; Conservative 44; Mismatches 66; Indels 3; Gaps 1;

OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;

SEQUENCE FROM N.A.

RP STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N., Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S., Horikoshi K.;
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001507; BAB03805.1; -
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS: TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 48.2%; Score 527; DB 16; Length 254;

Best Local Similarity 43.5%; Pred. No. 9.5e-37;

Matches 91; Conservative 57; Mismatches 61; Indels 0; Gaps 0;

QY 5 LVAVDIGNTSVNIGIFEKGLLANHLSVQAORMADEYASLLGLLQHAGIHPEELNRVI 64
DB 2 ILVLDVGNNTVLGVYQDETLLVHWRLATSRKTEDEYAMTVRSLEDFHAGLQFQDIDGIV 61
QY 65 MCSVVPPLTTTTEEVFKSYFKAAPLVVAGIKSGVKVRMDNPREVGADRIVNAARVLY 124
DB 62 ISSVVPNMFSLEQCKKYFHVTPMTIIGPICITGLNINIKYDNPKEVGADRIVNAARVLY 121
QY 125 PGACIIIVDMGTATFTDLSGGAYIGGAIAPGIATSAQAIAEKTSLPKRIEIRPAKVI 184
DB 122 GYPALVVDGTATFTCLINEKQYAGVIAPGLMISTEALYHRASKLPRIETAKPKQVVG 191
QY 185 SNTVSAMQSGIYFGYIGLVEELVRRIQTE 213
DB 182 TMTIDSMQSGIYFGYIGVQVGVVKKMAQ 210

RESULT 4

Q9F985 PRELIMINARY; PRT; 258 AA.
ID Q9F985;
AC Q9F985;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative 32 kDa replication protein.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;

SEQUENCE FROM N.A.

RP STRAIN=V;
RA Vasquez C., Pichantes S., Saavedra C.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF198621; AAG28531.1; -
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS: TIGR00671; baf; 1.
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 46.5%; Score 508; DB 2; Length 258;

Best Local Similarity 45.5%; Pred. No. 3.9e-35;

Matches 95; Conservative 46; Mismatches 68; Indels 0; Gaps 0;

QY 5 LVAVDIGNTSVNIGIFEKGLLANHLSVQAORMADEYASLLGLLQHAGIHPEELNRVI 64
DB 2 IFVLDVGNNTVLGVYDGLKHHRIETSRGKTEDEYGMTKALLNHVGLQFQSDIDGII 61

```

QY 65 MCSVVPPLTTTFFEEVFKSYKAAPLVVGAGIKSGVKVRMDNPREVGADRIYVAAAARVLY 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 ISSVVPPTFALEKMLKFIKPIVPGIKTGINKYDNPREVGADRIYVAVAGIHLV 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 PGACIIVDMGTATFTDLSEGGYVIGGAIAPGATSAQAIAEKTSKLPKIEIIRPAKVI 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 GSPLIIVDFGTATTCYINEHKQYMGGAIAPIGMISTEALPARAAKLPIRIEAPDDIIG 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 SNTVSAMQSGIYFGYIGVIEELVRRIOTE 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 KNTVSAMQAGLIGYVGQVEGIVSRMKAK 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
Q8XHL5 PRELIMINARY; PRT; 259 AA.
AC Q8XHL5
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB82174.1; -
DR InterPro; IPR004619; Baf.
DR TIGRFAMs; TIGR00515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 2819 MW; 0D5FEA3B7A145E10 CRC64;

```

```

Query Match 46.3%; Score 506.5; DB 16; Length 259;
Best Local Similarity 44.4%; Pred. No. 5.3e-35;
Matches 96; Conservative 55; Mismatches 62; Indels 3; Gaps 1;
QY 5 LVADIGNTSVNIGIFEGEKLANNHLSVAORMADEYASLLGLLQHGAIHP 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 ILLIDVGNITVGLTHDNKFIASWRISTDSKTSDEYSIQVMOLFNAKLNPEDVEGII 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 MCSVVPPLTTTFFEEVFKSYKAAPLVVGAGIKSGVKVRMDNPREVGADRIYVAAAARVLY 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 ISSVVPPTFALEKMLKFIKPIVPGIKTGINKYDNPREVGADRIYVAVAGIHLV 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 PGACIIVDMGTATFTDLSEGGYVIGGAIAPGATSAQAIAEKTSKLPKIEIIRPAKVI 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 KPMIIVDFGTATTCATEKGDLGNCICGQISADALFERAAKLPIRIEAPDDIIG 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 SNTVSAMQSGIYFGYIGVIEELVRRIOTE 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 KNTVSMQAGIYVIGRVEIVRKMCKEMMDLGEK 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
Q9X8N6 PRELIMINARY; PRT; 265 AA.
AC Q9X8N6
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.

```

```

GN SCO3380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL049628; CAB40880.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

```

```

Query Match 45.1%; Score 492.5; DB 16; Length 265;
Best Local Similarity 45.9%; Pred. No. 8.3e-34;
Matches 100; Conservative 42; Mismatches 65; Indels 11; Gaps 3;
QY 5 LVADIGNTSVNIGIFEGEKLANNHLSVAORMADEYASLLGLLQHGAIHP 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 ILLIDVGNITVGLTHDNKFIASWRISTDSKTSDEYSIQVMOLFNAKLNPEDVEGII 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 EELNRVIMCSVVPPLTTTFFEEVFKSYKAAPLV-VGAGIKSGVKVRMDNPREVGADRIYV 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 DGIDGICATVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSVPSV 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 AAAARVLPAGCIIVDMGTATFTDLSEGGYVIGGAIAPGATSAQAIAEKTSKLPKIEI 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 AVAAVELYGGPAAVVDFTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQRLKIEV 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 IRPAKVGISNTVSAMQSGIYFGYIGVIEELVRRIOTE 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 APRSVICKNTVEMQSGIYVGFAGQVGVNRMAREL 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

```

	Charbit A., Chetouani F., Couve E., de Daruvur A., Dehoux P., Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurge O., Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P., Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., Jones L.-M., Kerst U., Kreft J., Kuhn M., Kunst F., Kurapat G., Madueno E., Maifournam A., Mata Vicente J., Ng E., Nedjari H., Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R., Remmel B., Rose M., Schluter T., Simoes N., Tierrez A., Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.; "Comparative genomics of <i>Listeria species.</i> "; Science 294:849-852(2001)	
RL	EMBL; AL596164; CAC95486.1;	--
DR	Listinlist; LIN00253;	--
DR	InterPro; IPR004619; Baf.	
DR	Pfam; PF03309; Bvg_acc_factor; 1.	
DR	TIGRFAMS; TIGR00671; baf; 1.	
KW	Hypothetical protein; Complete proteome.	
DQ	SUBMITTER: 259 AA; 28227 MW; 554B03A0CEFA64F CRC64;	
SQ	SEQUENCE	44.2% Score 483; DB 16; Length 259; Best Local Similarity 40.7%; Pred. No. 5.2e-33; Matches 85; Conservative 55; Mismatches 69; Indels 0; Gaps 0
QY	5 LVAVDIGTSTVNTGIFGEKLLANWHLGSVQAORMADEVASLILGLQHAGIHPEELNRYI 64 :: : :: Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.	
Db	2 ILIVDVGNCTVTGVYKEQLLRHWRWTTHRHRTSDELGMTVLNLFVSAYNLTPSDIOGII 61 :: : ::	
QY	65 MCSVPPLTTTFEEVFYSKAAPLVVGCGIKSGVKVRMDNPREGADRIVNAAARVLV 124 : :: McShan D.E., et al.	
Db	62 ISSVVPIIMAMETCMCVNYRNIRLIIVGPGLKTGLNLVDMPREISGRINVAVASSEY 121 :: Issigmann R., et al.	
QY	125 PGACIIYDMGTATFTDLSEGGA YGGAIAPGIATSQAIAEAKTSKLPIEIIRPAKVIG 184 :: Petersen S.C., et al.	
Db	122 GTPVIYVDFGTAITFCYIDEGAVYGOGAIPAGIMSTEALYNRAAKLPFDVDAESSIQII 181 :: Gupta S.K., et al.	
QY	185 SNTVSAMQSIGIFYGYGLVELVRRIOTE 213 : :	
Db	182 KSTVASMOAGIFYGFIGOCEGIIAEKKQ 210 : :	
RESULT 9		
Q9A6Z1	ID Q9A6Z1 PRELIMINARY; PRU; 261 AA.	
AC	Q9A6Z1;	
DT	01-JUN-2001 (TrEMBRel. 17, Created)	
DT	01-JUN-2001 (TrEMBRel. 17, Last sequence update)	
DT	01-JUN-2002 (TrEMBRel. 21, Last annotation update)	
DE	Transcriptional activator, putative, Baf family. CC1935.	
OS	Caulobacter crescentus.	
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;	
OC	Caulobacter.	
RX	NCB1_Taxid=155892;	
RX	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ATCC 19089 / CB15;	
RX	MEDLINE=21173698; PubMed=11259647;	
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Nelson K.E., Niernan W.C., Reidblyum T.V., Laub M.T., Paulsen I.T., Paulson K.E., Potocnik R., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.; "Complete genome sequence of <i>Caulobacter crescentus.</i> "; Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).	
RL	EMBL; AE005867; AAC23910.1;	--
DR	TIGR; CC1935;	--
DR	InterPro; IPR004619; Baf.	
DR	Pfam; PF03309; Bvg_acc_factor; 1.	
DR	TIGRFAMS; TIGR00671; baf; 1.	
KW	Complete proteome.	
SQ	SEQUENCE 261 AA; 27965 MW; C19E60DB7B0714EF5 CRC64;	

Query Match 43.7%; Score 478; DB 16; Length 261;
 Best Local Similarity 44.9%; Pred. No. 1.4e-32;
 Matches 96; Conservative 41; Mismatches 75; Indels 2; Gaps 1;

QY 5 LVADIGNTSVNIGIFEGEKLANNHLSGVAORMADEYASLLGLLQHAGIHPEELNRYI 64
 DB 3 LLAIDVNTHTVGLSSGKEHAKVVQWRIRTESEVTADELALIDGLI---GDDSERL 58
 QY 65 MCSVVPPLTTTFEEVEKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRVNNAARVLY 124
 DB 63 ISSVVPQSFENLRNLSRYFNVEPLVIGENAKLGDIVRIKPESEAGADRLVNAIGAAVYI 122
 QY 125 PGACIIVDMGTRATTTFTLSEGGAYTGGAIAPGATNSAQAIKTSKLPKIEIRPA--KV 182
 DB 123 PGCLVIVDSGTATTEDIVAADCAFEFGIAPGCLNSMQALHEAAKLPRIAIORPAGNRI 182
 QY 183 ICSNTVSAMQSGIYFGYIGLVEELVRRRIQTQLGQ 216
 DB 183 VGTDTVSAMQSGVFWGYSISLIEGLVARIKAEKGE 216

RESULT 10

Q9CD56 PRELIMINARY; PRT; 274 AA.
 ID Q9CD56
 AC Q9CD56
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical protein ML0232.
 GN ML0232.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL583917; CAC29740.1; -.
 DR Leproma; ML0232; -.
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMs; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDE878765 CRC64;

Query Match 34.1%; Score 373; DB 16; Length 274;
 Best Local Similarity 35.8%; Pred. No. 1.1e-23;
 Matches 77; Conservative 52; Mismatches 78; Indels 8; Gaps 3;

QY 5 LVADIGNTSVNIGIFEGE---KLLANNHLSGVAORMADEYASLLGLLQHAGIHPEEL 60
 DB 2 LLAIDVNTHTVGLSSGKEHAKVVQWRIRTESEVTADELALIDGLI---GDDSERL 58
 QY 61 NRVMCSVVPPLTTTFEEVEKSYFKAAP-LVVGAGIKSGVKVRMDNPREVGADRVNNA 119
 DB 59 AGAALSTVPSVLHVRIMLQYWPSPVPHVLPICVGTGIPLLVNDNPREVGADRVNCL 118
 QY 120 ARVLYPGACIIVDMGTRATTTFTLSEGGAYTGGAIAPGATNSAQAIKTSKLPKIEIRP 179
 DB 119 AFHKGQAAIIVDFGSSICVDVWSAKGEFLGGAIPGVQVSSDAARSAALRRVELARP 178

QY 180 AKVIGNTVSAMQSGIYFGYIGLVEELVRRRIQTQL 214
 DB 179 RSVGKNVTECHQAGVWFGFAGLVGLVGRMRQDV 213

RESULT 11

O06282 PRELIMINARY; PRT; 272 AA.
 ID O06282
 AC O06282
 DT 01-JUL-1997 (TReMBLrel. 04, Created)
 DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative, Baf family).
 GN RV3600C OR MTCY07H7B.22 OR MT3706.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekai F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / OSHKOSH;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 295557; CAB08944.1; -.
 DR EMBL; AE007170; AAK48063.1; -.
 DR TIGR; MT3706; -.
 DR TubercuList; RV3600C; -.
 DR InterPro; IPR004619; Baf.
 DR Pfam; PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMs; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 272 AA; 29304 MW; 5D70E6EF0F09AC8B CRC64;

Query Match 33.9%; Score 371; DB 16; Length 272;
 Best Local Similarity 36.3%; Pred. No. 1.7e-23;
 Matches 78; Conservative 49; Mismatches 80; Indels 8; Gaps 3;

QY 5 LVADIGNTSVNIGIFEGE---EKLANNHLSGVAORMADEYASLLGLLQHAGIHPEEL 60
 DB 2 LLAIDVNTHTVGLSSGKEHAKVVQWRIRTESEVTADELALIDGLI---GEDSERL 58
 QY 61 NRVMCSVVPPLTTTFEEVEKSYFKAAP-LVVGAGIKSGVKVRMDNPREVGADRVNNA 119
 DB 59 TGAALSTVPSVLHVRIMLQYWPSPVPHVLPICVGTGIPLLVNDNPREVGADRVNCL 118
 QY 120 ARVLYPGACIIVDMGTRATTTFTLSEGGAYTGGAIAPGATNSAQAIKTSKLPKIEIRP 179
 DB 119 YDFRKAIIIVDFGSSICVDVWSAKGEFLGGAIPGVQVSSDAARSAALRRVELARP 178
 QY 180 AKVIGNTVSAMQSGIYFGYIGLVEELVRRRIQTQL 214

```

Db 179 RSVVGKNTVECMQAGAVFGAGLVGLVGRREDV 213
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
RESULT 12
032514 PRELIMINARY; PRT; 212 AA.
ID Q32514
AC Q32514;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to Bacillus subtilis.
OS Desulfovibrio vulgaris (strain Miyazaki).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MIYAZAKI;
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,
  Kohno K., Akutsu H., Kumagai I., Nakaya T.;
  "Sequence analyses of two ferredoxin genes and their flanking regions
  from Desulfovibrio vulgaris (Miyazaki F).";
  J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).
RL EMBL; AB005350; BAA21476.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
SQ SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;

Query Match 31.9%; Score 348.5; DB 2; Length 212;
Best Local Similarity 38.5%; Pred. No. 9.7e-22;
Matches 80; Conservative 39; Mismatches 84; Indels 5; Gaps 3;

QY 1 MSELKLVADIGNTSVNIGIFEGEKKLLANHLGSAQVMADEYASLLGLLQHAGIHPPEEL 60
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 1 MTQHFLFDIGNTVKIGTAVETAVLTSVLPDQCQTTDSIGRLLEVLHAGLGPADV 60
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 61 NRVMCSVVPPLTTTFEEVFKSYFKAAPLVGAGIKSGVKVMDNPREVGADRIVNAAAA 120
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 61 GACVASSVPGVGNPLIRACERYLYRKLFPAGDIAIPLDNRYERPAEYAGADRLVAAYAA 120
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 121 RVLYPG--ACIIIVDMGTATFTLSEGGAYIGGAIAPGATSAQAETSKLPKTI--EI 176
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 121 RRLYPGRSLVSVDFGTATFCV--EGGAYLGLGICPGVLSSAGALSSRTAKLPRISLEV 179
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 177 IRPAKVIKNTSVAMOSGIVFGVIGLVE 204
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 180 EEDSPVIGRSTTSLNHGIFGEAMTE 207
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|

RESULT 13
03RX54 PRELIMINARY; PRT; 262 AA.
ID Q9RX54
AC Q9RX54;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein DR0461.
GN DR0461.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
  Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
  Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
  Mathavan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
  Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
  Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

```

```

RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
  radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001905; AAF10040.1; -.
DR TIGR; DR0461; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 30.5%; Score 333.5; DB 16; Length 262;
Best Local Similarity 36.2%; Pred. No. 2.4e-20;
Matches 77; Conservative 36; Mismatches 95; Indels 5; Gaps 3;

QY 5 LVAVDIGNTSVNIGIFEGEKKLLAN--WHLGSAQVMADEYASLLGLLQHAGIHPPEELNRV 63
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 6 LLAVDIGNTTVLGLADASGALTHTWRTNREMLPDDLALQLHGLFTLAGAIP--RAA 63
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 64 IMCSVVPPLTTTFEEVFKSYFKAAPLVGAGIKSGVKVMDNPREVGADRIVNAARVL 123
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 64 VLSSVAPPYGVENYALAKRHFMDAFAPSAENLPDVTVELDTPGSGADRLCLNLFGEKY 123
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 124 YPGA--CIIVDMGTATFTLSEGGAYIGGAIAPGATSAQAETSKLPKTIIRPAK 181
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 124 LGLDYAVVDFGTSTNFDVVGRRFLGGLIATGAQVSADALFARAALPRITLQAPET 183
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 182 VIGSNTSVAMOSGIVFGVIGLVEELVRRIQTEL 214
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 184 AIGKNTVHALQSLVFGYAEWVDGLLRRAEL 216
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|

RESULT 14
Q8RFE4 PRELIMINARY; PRT; 256 AA.
ID Q8RFE4
AC Q8RFE4;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Bvg accessory factor.
GN FN0761.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
  Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
  Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Berni A.,
  Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
  Fongstein M., Kyripides N., Overbeek R.;
  "Genome sequence and analysis of the oral bacterium Fusobacterium
  nucleatum strain ATCC 25586";
  J. Bacteriol. 184:2005-2018(2002).
RL EMBL; AE010586; AAL94957.1; -.
KW Complete proteome.
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 30.2%; Score 330; DB 16; Length 256;
Best Local Similarity 32.1%; Pred. No. 4.5e-20;
Matches 71; Conservative 54; Mismatches 82; Indels 14; Gaps 5;

QY 5 LVAVDIGNTSVNIGIFEGE--KLLANHLGSAQVMADEYASLLGLLQHAGIHPPEELNRV 63
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
Db 2 IIGIDIGNTHVTGIVDNNGLISITFRIATNCKMTEDEYFSYFNNTKYNEISIKKVDAI 61
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 64 IMCSVVPPLTTTFEEVFKSYFKAAPLVGAGIK-----SGVKVMDNPREVGADRIVNA 117
      ||:| ||| ||:| ||:| ||:| ||:| ||:| ||:|

```

```
Db      62 LISSVVPNIIITOFFARKYFKVEATIVDLKLPFTFAKGI-----NYTGFADRIIDI 116
QY      118 AARVLYPGA-CIIIVDMGTATFTDLSEGGAYIGCAIAPGIATSAQAIAEKTSKLPKIEI 176
Db      117 TEAMQKTPDKNLVTFDGTATYFDVLRK-GVIIGGILPGIDMSINALYGTAKLPVKF 175
QY      177 IRPAKVGISNVSAMQSGIYFGYIGLVEELVRRIQTELQK 217
Db      176 TTPSSVLGTDIMKQIAIFGYAGQIKHIKKINEELNEE 216

RESULT 15
O83446
ID      O83446      PRELIMINARY;      PRU;      273 AA.
AC      O83446;
DT      01-NOV-1998 (TReMBLrel. 08, Created)
DT      01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT      01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE      Hypothetical protein TP0431.
OS      Treponema pallidum.
OC      Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX      NCBI_TaxID=160;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=NICHOLS;
RX      MEDLINE=98332770; PubMed=9665876;
RA      Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA      Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA      Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA      Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
RA      McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA      Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA      Venter J.C.;
RT      "Complete genome sequence of Treponema pallidum, the syphilis
RT      spirochete.";
RL      Science 281:375-388(1998).
DR      EMBL; AE001220; AAC65417.1; -.
DR      TIGR; TP0431; -.
DR      InterPro; IPR004619; Baf.
DR      Pfam; PF03309; Bvg_acc_factor; 1.
DR      TIGRFAMs; TIGR00671; baf; 1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match      28.4%; Score 310; DB 16; Length 273;
Best Local Similarity 33.0%; Pred. No. 2.4e-18;
Matches 72; Conservative 44; Mismatches 94; Indels 8; Gaps 3;
QY      5 LVAVDIGNTSVNIGIFEKEK-----LLANWHLSVAORMADEYASLLIGLQHAGIHPEE 59
Db      1 MLLIDVGNHVFGI-QGNGGRVCVRELFRLAPDARKTQDEYSLLIHAIACERAGVGRAS 59
QY      60 LNRVIMCSVPPLTFTTEEFKSYFKAAPLVVGGAGIKSGVKVRMDNP--REVGDRIVNA 117
Db      60 LRDAFISVVPLTKTIADAVAQISGVQPVVFGFWAYEHLFVRIPEVRAEIGTDLVANA 119
QY      118 AARVLYPGACIIIVDMGTATFTDLSEGGAYIGCAIAPGIATSAQAIAEKTSKLPKIEI 177
Db      120 VAAVYHFRSACVVVDCGTALTFTTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLA 179
QY      178 RPAKVGISNVSAMQSGIYFGYIGLVEELVRRIQTELQ 215
Db      180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELG 217

Search completed: June 24, 2003, 22:16:18
Job time : 20.2642 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 8.88544 Seconds
(without alignments)
725.188 Million cell updates/sec

Title: US-09-813-453A-57
Perfect score: 1093
Sequence: 1 MSEKLVAVDIGNTSVNIIGIF.....ICLVEELVRRIQTELQKTR 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	8.4	616	US-09-134-001C-3646	Sequence 3646, Ap
2	90.5	8.3	330	US-09-134-001C-3582	Sequence 3582, Ap
3	85.5	7.8	491	US-09-262-856A-5	Sequence 5, Appli
4	84	7.7	312	US-09-134-001C-3534	Sequence 3534, Ap
5	84	7.7	393	5229279-6	Patent No. 5229279
6	84	7.7	393	5512669-6	Patent No. 5512669
7	84	7.7	447	US-09-160-119-4	Sequence 4, Appli
8	84	7.7	674	US-09-160-119-2	Sequence 2, Appli
9	82.5	7.5	1529	US-09-134-001C-3945	Sequence 3945, Ap
10	78.5	7.2	507	US-09-091-097-34	Sequence 34, Appl
11	77.5	7.1	2183	US-08-348-891A-7	Sequence 7, Appli
12	77.5	7.1	2183	US-08-905-817-7	Sequence 7, Appli
13	77	7.0	1113	US-09-629-616-3	Sequence 3, Appli
14	76.5	7.0	366	US-09-134-001C-4365	Sequence 4365, Ap
15	76.5	7.0	947	US-09-228-986-73	Sequence 73, Appl
16	76	7.0	269	US-09-100-557-1	Sequence 12, Appl
17	76	7.0	359	US-09-347-798-12	Sequence 12, Appl
18	76	7.0	607	US-08-472-534-5	Sequence 5, Appli
19	75.5	6.9	686	US-08-849-480A-4	Sequence 4, Appli
20	75	6.9	485	US-08-818-112-70	Sequence 70, Appl
21	75	6.9	485	US-08-818-111-71	Sequence 71, Appl
22	75	6.9	485	US-09-056-556-70	Sequence 70, Appl
23	75	6.9	485	US-09-072-596-71	Sequence 71, Appl
24	75	6.9	683	US-09-213-293D-1	Sequence 1, Appli
25	75	6.9	889	PCT-US93-11725-2	Sequence 2, Appli
26	74.5	6.8	269	US-07-706-691G-1	Sequence 1, Appli
27	74.5	6.8	269	US-07-706-691G-2	Sequence 2, Appli

28	74.5	6.8	269	1	US-07-706-691G-3	Sequence 3, Appli
29	74.5	6.8	269	1	US-07-706-691G-6	Sequence 6, Appli
30	74.5	6.8	269	1	US-08-254-021-1	Sequence 1, Appli
31	74.5	6.8	269	1	US-08-254-021-2	Sequence 2, Appli
32	74.5	6.8	269	1	US-08-254-021-3	Sequence 3, Appli
33	74.5	6.8	269	1	US-08-254-021-6	Sequence 6, Appli
34	74.5	6.8	269	1	US-08-566-369-10	Sequence 10, Appl
35	74.5	6.8	269	1	US-08-566-369-11	Sequence 11, Appl
36	74.5	6.8	269	1	US-08-566-369-12	Sequence 12, Appl
37	74.5	6.8	269	1	US-08-566-369-13	Sequence 13, Appl
38	74.5	6.8	269	1	US-08-566-369-16	Sequence 16, Appl
39	74.5	6.8	269	1	US-08-566-369-17	Sequence 17, Appl
40	74.5	6.8	269	1	US-08-566-369-23	Sequence 23, Appl
41	74.5	6.8	269	1	US-08-566-369-24	Sequence 24, Appl
42	74.5	6.8	269	2	US-08-618-446-1	Sequence 1, Appli
43	74.5	6.8	269	2	US-08-618-446-2	Sequence 2, Appli
44	74.5	6.8	269	2	US-08-618-446-3	Sequence 3, Appli
45	74.5	6.8	269	2	US-08-618-446-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-3646
; Sequence 3646, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3646
; LENGTH: 616
; TYPE: FRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3646

Query Match	8.4%	Score 92;	DB 4;	Length 616;
Best Local Similarity	20.5%	Pred. No. 0.07;		
Matches	40;	Conservative 32;	Mismatches 67;	Indels 56; Gaps 7;
QY	4	KLVAVDIGNTSVNIIGIFEGE--KLLAN-----	-----WHLGSVAORMADEYAS 44	
Db	10	KVIGIDLTGTTSCVSLIEGDEPKVQNP	EGARTTSPVAFKNGETQVGEVAKKQATNPN 69	
QY	45	LLGLLQHQAGIH-----	-----PEELNRVIMCSVVPPLTTTFEEVFKSYFKAAPLVGGA 93	
Db	70	TVOSIKRHMGTDYKVDIEGKSVTPQELSAMILQN-----	-----LKSTAENYLGTDVTKAVITVPA 125	
QY	94	GKSGVKVRMDNPREGV---ADRIYNAAARVLYPG-----	-----ACIIYDMGTAT----- 137	
Db	126	YFNDGERQATKDAGKIAGLEVERIINPTAAALAYGLDKTETDQKVLVFDLGGGTPDVS 185		
QY	138	-----TFDTLSEGG 146		
Db	186	LELGDGVFEVLSTAG 200		

RESULT 2
US-09-134-001C-3582
; Sequence 3582, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

Query Match 7.7%; Score 84; DB 6; Length 393;
Best Local Similarity 23.5%; Pred. No. 0.3;
Matches 44; Conservative 25; Mismatches 66; Indels 52; Gaps 6;
QY 32 GSAQORMADEYASLL-GLLQAGHHPPELNVRVIMCSV- 68
DB 19 GSAKIPAPELGAVVYKAAALERAGVKPQVSEVIMQVLTAGSQNPARQAIAKAGLPAM 78
QY 69 VPPLTTTTEEVFKSYFKAAPLVVGVGAGIKSGVKVMDNPREVGADRVNAAARVLYPGAC 128
DB 79 VPAM--TINKVCGSLGKAVMLAANAIMAGDAEIVVAGGQE-----NMSAAPHVLPGRS 129
QY 129 IIVDMGTATTTDLTLEGG-----AYIGGAIAPGI-----ATSAQIAEKT 168
DB 130 DGFMRGDAKLVDTHIVDGLWDVYNYQYHMGITAENVVAKEYGITREAQDEFAVGSQNKAEAA 189
QY 169 SKLPKIE 175
DB 190 QKAGKFD 196
RESULT 6
5512669-6
Patent No. 5512669
APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
TITLE OF INVENTION: GENE ENCODING BACTERIAL ACETOACETYL-COA
REDUCTASE
NUMBER OF SEQUENCES: 7
CURRENT APPLICATION DATA:
APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
FILING DATE: 29-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 124,570
FILING DATE: 20-SEP-1993
APPLICATION NUMBER: 944,488
FILING DATE: 03-NOV-1992
APPLICATION NUMBER: 566,535
FILING DATE: 13-AUG-1990
APPLICATION NUMBER: 67,695
FILING DATE: 29-JUN-1987
SEQ ID NO: 6:
LENGTH: 393
5512669-6

Query Match 7.7%; Score 84; DB 6; Length 393;
Best Local Similarity 23.5%; Pred. No. 0.3;
Matches 44; Conservative 25; Mismatches 66; Indels 52; Gaps 6;
QY 32 GSAQORMADEYASLL-GLLQAGHHPPELNVRVIMCSV- 68
DB 19 GSAKIPAPELGAVVYKAAALERAGVKPQVSEVIMQVLTAGSQNPARQAIAKAGLPAM 78
QY 69 VPPLTTTTEEVFKSYFKAAPLVVGVGAGIKSGVKVMDNPREVGADRVNAAARVLYPGAC 128
DB 79 VPAM--TINKVCGSLGKAVMLAANAIMAGDAEIVVAGGQE-----NMSAAPHVLPGRS 129
QY 129 IIVDMGTATTTDLTLEGG-----AYIGGAIAPGI-----ATSAQIAEKT 168
DB 130 DGFMRGDAKLVDTHIVDGLWDVYNYQYHMGITAENVVAKEYGITREAQDEFAVGSQNKAEAA 189
QY 169 SKLPKIE 175
DB 190 QKAGKFD 196

RESULT 7
US-09-160-119-4
Sequence 4, Application US/09160119A
Patent No. 6316219
GENERAL INFORMATION:
APPLICANT: KRIEF, STEPHANE
APPLICANT: SOUCHET, MICHEL
APPLICANT: BRIL, ANTOINE

TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GH-30985
CURRENT APPLICATION NUMBER: US/09/160,119A
CURRENT FILING DATE: 1998-09-24
EARLIER APPLICATION NUMBER: EP 97402511.6
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: EP 98401655.0
EARLIER FILING DATE: 1998-07-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 447
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-160-119-4

Query Match 7.7%; Score 84; DB 4; Length 447;
Best Local Similarity 24.3%; Pred. No. 0.36;
Matches 62; Conservative 33; Mismatches 102; Indels 58; Gaps 12;
QY 6 VAVDIGNTSV-----NIGIFEGEKLLANHILGSAQORMADEYASLLGLL- 50
DB 109 VAGAVGATAVYPIDLVKTRMQNRSTGFSVGMELTKNSFDCPKKVLRYEGFFGLYRGLDP 168
QY 51 QHAGIHPPELNVRVIMCSVVPPLTTTTFEEVFKSYFKAAPL---VVGAGIKSGVKVMDNPR 107
DB 169 QLLGVAPPEKAIKLTVDNFV-----RDKFHKHKGDSVPLAAEILAGGCAGGSOVIFTNPL 221
QY 108 EVGADRVNA-----AAARY-----LYPG--ACIIVDMG-TATTTDLTLEGG 147
DB 222 EIVKIRLQVAGEITTPRVSALSVVRDLGFFGIYKGAACFLRDPFSAIYFPCYAHVKA 281
QY 148 YIG---GAIAGIATSAQIAEKTSLPKIEIIRPAKVIGSNVTSAMQSG--LYFCYIGL 202
DB 282 SFANEDQVSPGSLLAGAIA-----GMPAASLTPADVTKRLQVAARAGQTTYSGVI--- 335
QY 203 VEELVRRITQELGQK 217
DB 336 --DCFRKILREBGP 348

RESULT 8
US-09-160-119-2
Sequence 2, Application US/09160119A
Patent No. 6316219
GENERAL INFORMATION:
APPLICANT: KRIEF, STEPHANE
APPLICANT: SOUCHET, MICHEL
APPLICANT: BRIL, ANTOINE
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GH-30985
CURRENT APPLICATION NUMBER: US/09/160,119A
CURRENT FILING DATE: 1998-09-24
EARLIER APPLICATION NUMBER: EP 97402511.6
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: EP 98401655.0
EARLIER FILING DATE: 1998-07-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 674
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-160-119-2

Query Match 7.7%; Score 84; DB 4; Length 674;
Best Local Similarity 24.3%; Pred. No. 0.66;
Matches 62; Conservative 33; Mismatches 102; Indels 58; Gaps 12;
QY 6 VAVDIGNTSV-----NIGIFEGEKLLANHILGSAQORMADEYASLLGLL- 50
DB 336 VAGAVGATAVYPIDLVKTRMQNRSTGFSVGMELTKNSFDCPKKVLRYEGFFGLYRGLDP 395

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 12.1708 seconds
(without alignments)
1947.059 Million cell updates/sec

Title: US-09-813-453a-57
Perfect score: 1093
Sequence: 1 MSEKLVAVDIGNTSVNIIGIF.....ICLVEELVRRIQTELQKTR 219
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 41779 segs, 108206813 residues
Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published_Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1093	100.0	219	9	US-09-813-453a-57
2	558	51.1	255	9	US-09-813-453a-7
3	527	48.2	254	9	US-09-813-453a-47
4	527	48.2	262	9	US-09-813-453a-45
5	518	47.4	233	9	US-09-813-453a-17
6	515	47.1	258	9	US-09-813-453a-2
7	513	46.9	250	9	US-09-813-453a-3
8	513	46.9	258	9	US-09-813-453a-49
9	492.5	45.1	265	9	US-09-813-453a-4
10	489	44.7	256	9	US-09-813-453a-55
11	478	43.7	260	9	US-09-813-453a-51
12	394	36.0	258	9	US-09-813-453a-6
13	371	33.9	272	9	US-09-712-363-276
14	371	33.9	272	9	US-09-813-453a-5
15	348.5	31.9	212	9	US-09-813-453a-59
16	333.5	30.5	262	9	US-09-813-453a-8
17	310	28.4	273	9	US-09-813-453a-10
18	279.5	25.6	246	9	US-09-813-453a-9
19	265.5	24.3	257	9	US-09-813-453a-53

20	190	17.4	262	9	US-09-813-453a-11	Sequence 11, Appl
21	189	17.3	241	9	US-09-813-453a-63	Sequence 63, Appl
22	184	16.8	244	9	US-09-813-453a-41	Sequence 41, Appl
23	153	14.0	460	9	US-09-813-453a-39	Sequence 39, Appl
24	152	13.9	592	9	US-09-813-453a-22	Sequence 22, Appl
25	150	13.7	592	9	US-09-813-453a-43	Sequence 43, Appl
26	144	13.2	257	9	US-09-813-453a-13	Sequence 13, Appl
27	143.5	13.1	242	9	US-09-813-453a-65	Sequence 65, Appl
28	136	12.4	248	9	US-09-813-453a-20	Sequence 20, Appl
29	134.5	12.3	249	9	US-09-813-453a-70	Sequence 70, Appl
30	134	12.3	249	9	US-09-813-453a-61	Sequence 61, Appl
31	130.5	11.9	267	9	US-09-813-453a-15	Sequence 15, Appl
32	130	11.9	229	9	US-09-813-453a-12	Sequence 12, Appl
33	108	9.9	223	9	US-09-895-913a-74	Sequence 74, Appl
34	108	9.9	223	9	US-09-813-453a-14	Sequence 14, Appl
35	108	9.9	223	9	US-09-813-453a-67	Sequence 67, Appl
36	88	8.1	509	10	US-09-731-872-389	Sequence 389, Appl
37	86.5	7.9	209	9	US-09-813-453a-21	Sequence 21, Appl
38	86	7.9	641	9	US-09-364-847-19	Sequence 19, Appl
39	85.5	7.8	491	10	US-09-987-190-5	Sequence 5, Appl
40	84.5	7.7	274	10	US-09-815-242-10868	Sequence 10868, A
41	84.5	7.7	2665	10	US-09-864-761-34248	Sequence 34248, A
42	84	7.7	393	9	US-09-364-847-2	Sequence 2, Appl
43	84	7.7	641	9	US-09-364-847-10	Sequence 10, Appl
44	32.5	7.5	323	9	US-10-116-255-61	Sequence 61, Appl
45	81	7.4	320	9	US-10-116-255-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-09-813-453a-57
; Sequence 57, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Dehalococcoides ethenogenes
US-09-813-453a-57

Query Match	100.0%;	Score	1093;	DB	9;	Length	219;
Best Local Similarity	100.0%;	Pred. No.	1.4e-97;				
Matches	219;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MSEKLVAVDIGNTSVNIIGIFEGEKKLLANNHLSVQAORMADEYASLLGLLQHHAGIHPPEL	60				
Db	1	MSEKLVAVDIGNTSVNIIGIFEGEKKLLANNHLSVQAORMADEYASLLGLLQHHAGIHPPEL	60				
QY	61	NRVIMCSVVPPLTTTTFEEVFKSYFKAAPLVVAGIKSGYKVRMDNPREVGADRVNAAA	120				
Db	61	NRVIMCSVVPPLTTTTFEEVFKSYFKAAPLVVAGIKSGYKVRMDNPREVGADRVNAAA	120				
QY	121	RVLYPGACIIIVDMGTATFTDITLSEGGAYIGGAIAPIGATSAQAIAEKTSLKPIEIRPA	180				
Db	121	RVLYPGACIIIVDMGTATFTDITLSEGGAYIGGAIAPIGATSAQAIAEKTSLKPIEIRPA	180				
QY	181	KVIGNTSVSAMSGIYFGYIGLVEELVRRIQTELQKTR	219				
Db	181	KVIGNTSVSAMSGIYFGYIGLVEELVRRIQTELQKTR	219				

Db 181 KVICNTVSAMQSGIYFGYIGLVVEELVRRITQELGOKTR 219

RESULT 2

US-09-813-453A-7

; Sequence 7, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OG2-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 7

LENGTH: 255

; TYPE: PRT

; ORGANISM: Geobacter sulfurreducens

US-09-813-453A-7

Query Match 51.1%; Score 558; DB 9; Length 255;

Best Local Similarity 49.8%; Pred. No. 5.8e-46;

Matches 104; Conservative 47; Mismatches 58; Indels 0; Gaps 0;

QY 5 LVAVDIGNTSVNIIGIFEKGLLANHVLGSAQMADEYASLLGLLQHAGIHPPEELNRVI 64

DB 2 LVAVDIGNTSVNIIGIFEKGLLANHVLGSAQMADEYASLLGLLQHAGIHPPEELNRVI 61

QY 65 MCSVVPPLTTTFFEEVFKSYFKAAPLVVVGAGIKSGVKVMDNPREVGADRVNNAARVLY 124

DB 62 ISSVVPPLTGLVRLSLGIFGMRPLVVGPGIKTGMPIQYDNPREVGADRVNNAARVLY 121

QY 125 PGACIIVDMGTATFTLSEGGAYIGGAIAPGATSAQIAETSKLPKIEIIRPAKVIG 184

DB 122 RTSLIIVDFGTATTCYINEKHMGVITPGLMISREALYSRAAKLPRIETKPSVVG 181

QY 185 SNTVSAMQSGIYFGYIGLVVEELVRRITQTE 213

DB 182 RNTVSNQAGIYGYVGLVDEIVTRKAE 210

RESULT 3

US-09-813-453A-47

; Sequence 47, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OG2-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 47

LENGTH: 254

; TYPE: PRT

; ORGANISM: Bacillus halodurans

US-09-813-453A-47

Query Match 48.2%; Score 527; DB 9; Length 254;

Best Local Similarity 43.5%; Pred. No. 5.6e-43;

Matches 91; Conservative 57; Mismatches 61; Indels 0; Gaps 0;

QY 5 LVAVDIGNTSVNIIGIFEKGLLANHVLGSAQMADEYASLLGLLQHAGIHPPEELNRVI 64

DB 2 ILVIDVGNNTVLGVYQDETLLVHHWRLATSRQKTEDEYAMTVRSRDLHAGLQPDIGIV 61

QY 65 MCSVVPPLTTTFFEEVFKSYFKAAPLVVVGAGIKSGVKVMDNPREVGADRVNNAARVLY 124

DB 62 ISVVPFPMFSLQDMCKKYFHVTPMIIIGPGIKTGLNIDPNKREVGADRVNNAARVLY 121

QY 125 PGACIIVDMGTATFTLSEGGAYIGGAIAPGATSAQIAETSKLPKIEIIRPAKVIG 184

DB 122 GYPAIVVDFGTATTCYINEKHMGVITPGLMISREALYSRAAKLPRIETKPSVVG 181

QY 185 SNTVSAMQSGIYFGYIGLVVEELVRRITQTE 213

DB 182 TNTIDSMQSGIFGYVGSQVDGVVVRKMAQ 210

RESULT 4

US-09-813-453A-45

; Sequence 45, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OG2-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 45

LENGTH: 262

; TYPE: PRT

; ORGANISM: Bacillus anthracis

US-09-813-453A-45

Query Match 48.2%; Score 527; DB 9; Length 262;

Best Local Similarity 47.0%; Pred. No. 5.8e-43;

Matches 101; Conservative 46; Mismatches 68; Indels 0; Gaps 0;

QY 5 LVAVDIGNTSVNIIGIFEKGLLANHVLGSAQMADEYASLLGLLQHAGIHPPEELNRVI 64

DB 2 IFVLDVGNNTVLGVFEEGELRQHWMETDRHKTEDEYGMVYKQLLEHGLSFDVKGII 61

QY 65 MCSVVPPLTTTFFEEVFKSYFKAAPLVVVGAGIKSGVKVMDNPREVGADRVNNAARVLY 124

DB 62 VSSVVPPIFALERMCEYFKIKPLVVGPGIKTGLNIDPNKREVGADRVNNAARVLY 121

QY 125 PGACIIVDMGTATFTLSEGGAYIGGAIAPGATSAQIAETSKLPKIEIIRPAKVIG 184

DB 122 GSPLIIVDFGTATTCYINEKHMGVITPGLMISREALYSRAAKLPRIETKPSVVG 181

QY 185 SNTVSAMQSGIYFGYIGLVVEELVRRITQELGOKTR 219

DB 182 KNTVSAMQSGIYGYVGVQVGVIGVVKRMEKAEKQEPK 216

RESULT 5

US-09-813-453A-17

; Sequence 17, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

	Query Match	47.1%	Score 515;	DB 9;	Length 258;	
	Best Local Similarity	45.8%;	Pred. No. 8.2e-42;			
	Matches	97;	Conservative 45;	Mismatches 70;	Indels 0;	Gaps 0;
<hr/>						
QY	5	LVAVDGNTGVNIGIFPEGEKLLANWHLGSVQAQRMADEYASLLGLLQHAGIHPEELNRVI	64			
		I : : : I I I : : : I I : : : I I : : : I I : : : I I : : : I I : : : I I :				
Dd	2	LLVIDVGNTVLGVGHDXGLEHWRITETSRHKTEDEFGMILSLFDHSGLMFQEIDGII	61			
<hr/>						
QY	65	MCSVVPPLTTTFEEVKFSYFKAAPLVYGAGIKSCVKVRMDNPREVGDADRVNAARVLY	124			
		: : : I I I I I : : : I I : : : I I : : : I I : : : I I : : : I I :				
Dd	62	ISSVWPPIMPALERMCTGYFHIEPQIVGPNGMKTGLNIKYKNPKVEGDADRVNAVAAIHL	121			

Query Match	46.9%	Score 513;	DB 9;	Length 250;
Best Local Similarity	45.6%	Pred. No. 1.2e-41;		
Matches 99;	Conservative 47;	Mismatches 69;	Indels 2;	Gaps 1

RESULT 8
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1

```

> APPLICANT: PATERSON, THOMAS A.
> TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
> TITLE OF INVENTION: ANTIBIOTICS
> FILE REFERENCE: OZG-001
> CURRENT APPLICATION NUMBER: US/09/813,453A
> CURRENT FILING DATE: 2001-03-20
> PRIOR APPLICATION NUMBER: US 60/227,860
> PRIOR FILING DATE: 2000-08-24
> PRIOR APPLICATION NUMBER: 09/667,569
> PRIOR FILING DATE: 2000-09-21
> NUMBER OF SEQ ID NOS: 77
> SOFTWARE: PatentIn Ver. 2.0
> SEQ ID NO 49

```


Db 2 LLAIEQGNNTMFIAIHGASVQAQWRSATESRTRADEYVVMWSQLLSMOGLGFRDAIV 61
Qy 65 MCSVVPPLTTTFFEEVFKSYKAAPLVVGAGKSGVKVMDNPREVGADRIVNAARVLY 124
Db 62 ISSVVPQSFNRLNLSRRYFNEPLVIGENAKLIDVRIEKPSSAGADRLVNAICAAVY 121
Qy 125 PGACIIVDMGPTATFTDLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIRPA--KV 182
Db 122 PGLVWIDSGTATFTDIVAADGAFEGGIAPGINLSMOALHEAAAKLPRIAIQRPAGNRI 181
Qy 183 IGSNTVSAMQSGYFYGLVGLVEELVRRRIOTELGQ 216
Db 182 VGTDTVSAMQSGVFWGYLSIEGLVLARIKAERGE 215

RESULT 12

US-09-813-453A-6
Sequence 6, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813.453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 6
LENGTH: 258
TYPE: PRT
ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

Query Match 36.0%; Score 394; DB 9; Length 258;
Best Local Similarity 40.0%; Pred. No. 3.7e-30;
Matches 84; Conservative 39; Mismatches 85; Indels 2; Gaps 2;
Qy 5 LVAVDIGNTSVNIIGIFEKLLANHLGSAQRMADYASLLGLLQHGAGIHPEELNRY 64
Db 2 LLCIDCGNTNTVSWDGTDFATWRIATDHRRTADEYFVWLTLMLQLGLQ-GRISEAI 60
Qy 65 MCSVVPPLTTTFFEEVFKSYKAAPLVVG-AGIKSGVKVMDNPREVGADRIVNAARVLY 123
Db 61 ISSAPRVVFNRLVLCNRYFDCRPYVVGKPCGELPVPAPRVDPGTGTGPDRLVNTVAGYDR 120
Qy 124 YPGACIIVDMGPTATFTDLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIRPAKVI 183
Db 121 HGGDLIVDFGTATFTDVAADGAFEGGIAPGINLSMOALHEAAAKLPRIAIQRPAGNRI 180
Qy 184 GSNVTSAMQSGYFYGLVGLVEELVRRRIOTE 213
Db 181 GTNVACIQSGVWGYGLVGSIVRIQIRME 210

RESULT 13

US-09-712-363-276
Sequence 276, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712.363
CURRENT FILING DATE: 2000-11-13

PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206,
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 276
LENGTH: 272
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-276

Query Match 33.9%; Score 371; DB 9; Length 272;
Best Local Similarity 36.3%; Pred. No. 6.5e-28;
Matches 78; Conservative 49; Mismatches 80; Indels 8; Gaps 3;
Qy 5 LVAVDIGNTSVNIIGIFEKLLANHLGSAQRMADYASLLGLLQHGAGIHPEEL 60
Db 2 LLAIDVRNTHTVGLLSGMKEHAKVYQWRIETSEVTADELALITDGLI---GEDSERL 58
Qy 61 NRVIMCSVVPPLTTTFFEEVFKSYKAAAP-LVVGAGIKSGVKVMDNPREVGADRIVNAAR 119
Db 59 TGTAAALSTVPSVLHEVRIMLDQYWPSPVPHVLPVGTGIPLLVDNPKVEVGADRIVNCIA 118
Qy 120 ARVLYPGACIIVDMGPTATFTDLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIRP 179
Db 119 AYDRFKAAIIVDFGSSICVDVWSAKGFEGLGAIAPGVQVSSDAANAARSALRRVELARP 178
Qy 180 AKVIGSNTVSAMQSGYFYGLVGLVEELVRRRIOTEL 214
Db 179 RSVGKNTVECMQACAGAVEGAGLVGLVGRIREDV 213

RESULT 14

US-09-813-453A-5
Sequence 5, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OG2-001
CURRENT APPLICATION NUMBER: US/09/813.453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 272
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5
Query Match 33.9%; Score 371; DB 9; Length 272;
Best Local Similarity 36.3%; Pred. No. 6.5e-28;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: June 24, 2003, 22:03:27 ; Search time 9.03478 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453A-57
Perfect score: 1093
Sequence: 1 MSEKLVAVDIGNTSVNIQIF.....IGLVEELVRRIQTELQKTR 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	533.5	48.8	273	2 E97293	probable transcrip
2	527	48.2	254	2 F83660	hypothetical prote
3	518	47.4	233	2 S66100	conserved hypotet
4	492.5	45.1	265	2 T36391	hypothetical prote
5	489	44.7	259	2 AF1102	conserved hypotet
6	483	44.2	259	2 AF1464	conserved hypotet
7	478	43.7	261	2 B87489	transcription acti
8	373	34.1	274	2 H86937	conserved hypotet
9	371	33.9	272	2 A70955	hypothetical prote
10	333.5	30.5	262	2 E75516	conserved hypotet
11	310	28.4	273	2 D71326	conserved hypotet
12	279.5	25.6	246	2 D72320	conserved hypotet
13	190	17.4	262	2 F70165	conserved hypotet
14	160	14.6	276	2 A12292	hypothetical prote
15	152	13.9	592	2 H82031	probable biotin-la
16	150	13.7	592	2 B81009	BirA protein/Bvg a
17	144	13.2	257	2 S75559	hypothetical prote
18	143.5	13.1	242	2 A82637	conserved hypotet
19	136	12.4	248	2 H83111	hypothetical prote
20	130.5	11.9	267	2 T40327	baf protein - Bord
21	130	11.9	229	2 E70465	hypothetical prote
22	118	10.8	223	2 G71887	hypothetical prote
23	108	9.9	223	2 F64627	hypothetical prote
24	94	8.6	253	2 C70317	cyclase hisf [simi
25	91	8.3	427	2 D64353	UDP-N-acetyl-D-man
26	91	8.3	453	2 D82140	C4-dicarboxylate t
27	89.5	8.2	396	2 AH3611	acetyl-CoA C-acety
28	89.5	8.2	420	2 B82081	cell division prot
29	89	8.1	420	2 AC3284	D-amino-acid dehyd

30	88.5	8.1	559	2 B75477	conserved hypotet
31	88.5	8.1	560	2 AB2043	chaperonin GroEL [
32	88	8.1	394	2 S29276	acetyl-CoA C-acety
33	88	8.1	404	2 C46665	aspartate kinase (
34	87.5	8.0	328	1 S23352	glucose kinase - S
35	87.5	8.0	484	2 B64481	hypothetical prote
36	87.5	8.0	603	2 T05416	probable phosphogl
37	87	8.0	418	2 AD0069	cell division prot
38	87	8.0	428	2 A97341	adenylosuccinate s
39	86.5	7.9	209	2 H81382	hypothetical prote
40	86.5	7.9	429	2 E75107	thiamin biosynthes
41	86	7.9	837	2 S30971	gene 26 protein -
42	85.5	7.8	492	2 F98125	IMP dehydrogenase
43	85.5	7.8	492	2 C95260	IMP dehydrogenase
44	84.5	7.7	938	2 AB1530	transcription regu
45	84.5	7.7	938	2 A11172	transcription regu

ALIGNMENTS

RESULT 1

E97293

probable transcription regulator, homolog of Bvg accessory factor [Imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97293
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97293
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-273 <KUR>

A:Cross-references: GB:AB001437; PIDN:AAK81136.1; PID:gl5026270; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3200

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	48.8%	Score	533.5	DB 2	Length	273	
Best Local Similarity	47.7%	Pred. No.	7.6e-37				
Matches	103	Conservative	44	Mismatches	66	Gaps	1
QY	5	LVAVDIGNTSVNIQIFEGEKKLANVHLGSAVQMADEYASLLGLLQVAGIHPPEELNRVI	64				
Db	13	ILVLDVGNINVLGIYNDTKLTAEWRLSTDVLSADEYGIQVMNLFQDDKLDPTLVEGVI	72				
QY	65	MCSVPEPLTTTFEEVFKSYFKAAPLVGAGIKSGVKVMDNPREGVQADRVNAAAAARVLY	124				
Db	73	ISSVVPNIMVLSLEHMIRKYNINPLVGVGPKTGINKYDNPKEVQADRVNAAVAHEIY	132				
QY	125	PGACIIVDMGATTTDTLSEGGAYIGGATAPGIATSAQATAERTSKLPKIEIRPAKVIG	184				
Db	133	KRSLIIDFGTATTCVAVRENGDYLGGAICPGIKVSSSEALFEKAKLPVELIKPAYAIC	192				
QY	185	SNTVSAMQSGIYFGYIGLVVELVRRITQTEL---GQK	217				
Db	193	KNTISSIOSGIYVYIGQVRYIVERMKKELOEERGEK	228				

RESULT 2

F83660

hypothetical protein BH0086 [Imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: F83660
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: F83660
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <STO>
A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:GN000004
A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 48.2%; Score 527; DB 2; Length 254;
Best Local Similarity 43.5%; Pred. No. 2.4e-36;
Matches 91; Conservative 57; Mismatches 61; Indels 0; Gaps 0;

QY 5 LVAVDIGNTSVNIIGFEGEKKLANHLSVGAORMADEYASLLGLLQHAGIHPPEELNRYI 64
DB 2 ILVIDGNTNTVLGVYQDETIVHWRLATSRQKTEDEYAMTVRSFDFHAGLQFQDIDGIV 61

QY 65 MCSVVPPLTTTTEEFKSYFKAAPLVVAGIKSGVKVRMDNPREVGADRVNNAARVLY 124
DB 62 ISSVPPMFLSEQMKKFFHYTPMTIIGPGIKTGLNPKYDNPKEVGADRVNNAARVLY 121

QY 125 PGACIIVDMGTATFTDLSEGGAYIGGAIAPAGIATSAQAIAEKTSLPKRIETIRPAKVIG 184
DB 122 GPVAVVDGTATYCLINEKKQYAGGVITAPGLIMISTEALYHRASKLPRIETAKPKQVVG 181

QY 185 SNTVSAMQSGIFGYIGLVEELVRRRIQTE 213
DB 182 TWTIDSMQSGIFGYIGVQVGVKRMKAQ 210

RESULT 3
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C:Accession: S66100; E69740
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chromosome 390, 249-256, 1997
A:Reference number: S65967; MUID:96051385; PMID:7584024
A:Accession: S66100
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <OGA>
A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BA05305.1; PID:d1005847; PID:g467459
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R:Kunst, F.; Ogasawara, N.; Woszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
T.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
Y, M.; Ogasawa, K.; Ogawa, B.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69740
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <KUN>
A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:el182003;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yacB
A:Start codon: TTG
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 47.4%; Score 518; DB 2; Length 233;
Best Local Similarity 45.6%; Pred. No. 1.2e-35;
Matches 98; Conservative 45; Mismatches 72; Indels 0; Gaps 0;

QY 5 LVAVDIGNTSVNIIGFEGEKKLANHLSVGAORMADEYASLLGLLQHAGIHPPEELNRYI 64
DB 2 LLVIDGNTNTVLGVYHDKLEYHWRITSRKTEDEFGMLRSFDFHAGLQFQDIDGII 61

QY 65 MCSVVPPLTTTTEEFKSYFKAAPLVVAGIKSGVKVRMDNPREVGADRVNNAARVLY 124
DB 62 ISSVPPMFLSEQMKKFFHYTPMTIIGPGIKTGLNPKYDNPKEVGADRVNNAARVLY 121

QY 125 PGACIIVDMGTATFTDLSEGGAYIGGAIAPAGIATSAQAIAEKTSLPKRIETIRPAKVIG 184
DB 122 GNPLIVDGTATYCYIDENQYMGAIAPAGITISTEALYSRAAKLPRIETIRPDNIIG 181

QY 185 SNTVSAMQSGIFGYIGLVEELVRRRIQTELGOKTR 219
DB 182 KNTVSAMQSGILFGYVQGVGVKRMKAQKQDPR 216

RESULT 4
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36391
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T36391
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-265 <OLJ>
A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE94.31c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 45.1%; Score 492.5; DB 2; Length 265;
Best Local Similarity 45.9%; Pred. No. 1.8e-33;
Matches 100; Conservative 42; Mismatches 65; Indels 11; Gaps 3;

QY 5 LVAVDIGNTSVNIIGFEGEKKLANHLSVGAORMADEYASLLGLLQHAGIHP----- 57
DB 2 LITIDVGNTHTVLGLFDGEDIEVHWRISTDSRTADELAVLQGLM---GMPLLGDELG 58

QY 58 EELNRVIMGSVVPPLTTTTEEFKSYFKAAPLVVAGIKSGVKVRMDNPREVGADRVN 116
DB 59 DGIDGIAICATVPSVLHELREVTTRYGVDPAVLVEPGVTKVPILTDPHKEVGADRIIN 118

QY 117 AAAARVLPACIIVDMGTATFTDLSEGGAYIGGAIAPAGIATSAQAIAEKTSLPKRIE 176
DB 119 AVAAVLEGGPVAIVDFGTATFADVARGEYIGGVIGVTSVEALGVKAQLRKIEV 178

QY 177 IRPAKVISGTSVMSQSGIFGYIGLVEELVRRRIQTEL 214
DB 179 APRSVIGKNTVEAMQSGIYVGFAGQVDGVNVMAREL 216

RESULT 5
AF1102
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EG
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AF1102
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.;

QY 5 LVAVDIGNTSVNIIGFEK-----KLLANNHLGSAQMADEYASLLGLLQHGAIHPEEL 60
 Db 2 LLAVDIGNTSVNIIGFEK-----KLLANNHLGSAQMADEYASLLGLLQHGAIHPEEL 58
 QY 61 NRVMCSVPPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 119
 Db 59 AGAALSTVPSVLHVRIMLDQWPSVPHVLEPVGRTGPIPLVDNPKVGVADRIYVNC 118
 QY 120 ARVLYPCACIIVDMGTATFTDLSEGGAVIGGAIAPGIATSAQAIAEKTSKLPKIEIIRP 179
 Db 119 AFHKFQAAIVVDGSSICVDVVSAGKEFLGAIAPGVQVSSDAAARSALRRVELARP 178
 QY 180 AKVIGNTVSAMQSGIYFGYIGLVEELVRIOTEL 214
 Db 179 RSVVGKNTVECMQAGVFGFAGLVGLVGRMRQDV 213

RESULT 9

A70955

Query Match 33.9%; Score 371; DB 2; Length 272;
 Best Local Similarity 36.3%; Pred. No. 2.2e-23;
 Matches 78; Conservative 49; Mismatches 80; Indels 8; Gaps 3;
 QY 5 LVAVDIGNTSVNIIGFEK-----EKLANNHLGSAQMADEYASLLGLLQHGAIHPEEL 60
 Db 2 LLAVDIGNTSVNIIGFEK-----EKLANNHLGSAQMADEYASLLGLLQHGAIHPEEL 58
 QY 61 NRVMCSVPPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 119
 Db 59 AGAALSTVPSVLHVRIMLDQWPSVPHVLEPVGRTGPIPLVDNPKVGVADRIYVNC 118
 QY 120 ARVLYPCACIIVDMGTATFTDLSEGGAVIGGAIAPGIATSAQAIAEKTSKLPKIEIIRP 179
 Db 119 AFHKFQAAIVVDGSSICVDVVSAGKEFLGAIAPGVQVSSDAAARSALRRVELARP 178
 QY 180 AKVIGNTVSAMQSGIYFGYIGLVEELVRIOTEL 214
 Db 179 RSVVGKNTVECMQAGVFGFAGLVGLVGRMRQDV 213

RESULT 10

E75516

Query Match 28.4%; Score 310; DB 2; Length 273;
 Best Local Similarity 33.0%; Pred. No. 2.5e-18;
 Matches 72; Conservative 44; Mismatches 94; Indels 8; Gaps 3;
 QY 5 LVAVDIGNTSVNIIGFEK-----LLANNHLGSAQMADEYASLLGLLQHGAIHPEE 59
 Db 1 LLAVDIGNTSVNIIGFEK-----LLANNHLGSAQMADEYASLLGLLQHGAIHPEE 59
 QY 60 LNRVMCSVPPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 117
 Db 60 LRDATISSVPPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 119
 QY 118 AAARVLYPCACIIVDMGTATFTDLSEGGAVIGGAIAPGIATSAQAIAEKTSKLPKIEI 177
 Db 120 VAAVHFRSACVVDGCTALTFTAVDGTGLIGVVAIPGLRTAVQSLHTGTALQPLVPLA 179
 QY 178 RPAKVGNTVSAMQSGIYFGYIGLVEELVRIOTEL 215

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: E75516

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-262 <WHI>

A:Cross-references: GB:AE001905; GB:AE005513; NID:g6458144; PIDN:AAF10040.1; PID:g645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0461

A:Map position: 1

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.5%; Score 333.5; DB 2; Length 262;

Best Local Similarity 36.2%; Pred. No. 2.7e-20;

Matches 77; Conservative 36; Mismatches 95; Indels 5; Gaps 3;

QY 5 LVAVDIGNTSVNIIGFEK-----KLLANNHLGSAQMADEYASLLGLLQHGAIHPEELNRV 63

Db 6 LLAVDIGNTSVNIIGFEK-----KLLANNHLGSAQMADEYASLLGLLQHGAIHPEELNRV 63

QY 64 IMCVSVPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 123

Db 64 VLSSVAPPVGENVALAKRHFMDIDAFVSAENLPDVTVELDTPGSGVADRLCLNFGAEKY 123

QY 124 YPGA--CIIVDMGTATFTDLSEGGAVIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAK 181

Db 124 LGGLDYAVVVDGCTALTFTAVDGTGLIGVVAIPGLRTAVQSLHTGTALQPLVPLA 183

QY 182 VIGSNTVSAMQSGIYFGYIGLVEELVRIOTEL 214

Db 184 AIGKNTVHALQSLGVEGYAEMVDGLLRIRAE 216

RESULT 11

D71326

Query Match 33.9%; Score 371; DB 2; Length 272;
 Best Local Similarity 36.3%; Pred. No. 2.2e-23;
 Matches 78; Conservative 49; Mismatches 80; Indels 8; Gaps 3;
 QY 5 LVAVDIGNTSVNIIGFEK-----EKLANNHLGSAQMADEYASLLGLLQHGAIHPEEL 60
 Db 2 LLAVDIGNTSVNIIGFEK-----EKLANNHLGSAQMADEYASLLGLLQHGAIHPEEL 58
 QY 61 NRVMCSVPPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 119
 Db 59 AGAALSTVPSVLHVRIMLDQWPSVPHVLEPVGRTGPIPLVDNPKVGVADRIYVNC 118
 QY 120 ARVLYPCACIIVDMGTATFTDLSEGGAVIGGAIAPGIATSAQAIAEKTSKLPKIEIIRP 179
 Db 119 AFHKFQAAIVVDGSSICVDVVSAGKEFLGAIAPGVQVSSDAAARSALRRVELARP 178
 QY 180 AKVIGNTVSAMQSGIYFGYIGLVEELVRIOTEL 214
 Db 179 RSVVGKNTVECMQAGVFGFAGLVGLVGRMRQDV 213

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: D71326

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <COL>

A:Cross-references: GB:AE001220; GB:AE005520; NID:g3322705; PIDN:AAC65417.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0431

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.4%; Score 310; DB 2; Length 273;

Best Local Similarity 33.0%; Pred. No. 2.5e-18;

Matches 72; Conservative 44; Mismatches 94; Indels 8; Gaps 3;

QY 5 LVAVDIGNTSVNIIGFEK-----LLANNHLGSAQMADEYASLLGLLQHGAIHPEE 59

Db 1 LLAVDIGNTSVNIIGFEK-----LLANNHLGSAQMADEYASLLGLLQHGAIHPEE 59

QY 60 LNRVMCSVPPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 117

Db 60 LRDATISSVPPPLTTTTEEVFKSYFKAAP-LVVGAGIKSGVKVMDNPREVGADRIYNA 119

QY 118 AAARVLYPCACIIVDMGTATFTDLSEGGAVIGGAIAPGIATSAQAIAEKTSKLPKIEI 177

Db 120 VAAVHFRSACVVDGCTALTFTAVDGTGLIGVVAIPGLRTAVQSLHTGTALQPLVPLA 179

QY 178 RPAKVGNTVSAMQSGIYFGYIGLVEELVRIOTEL 215

Db 180 LPDSVLGKDTTHAVQAGVGRGTFLFVIRAMIAQCOKELG 217

RESULT 12
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: D72320
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: D72320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-246 <ARN>
A:Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g4981417
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0883
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 25.6%; Score 279.5; DB 2; Length 246;
Best Local Similarity 37.6%; Pred. No. 7.4e-16;
Matches 79; Conservative 30; Mismatches 92; Indels 9; Gaps 6;

Qy 8 VDIGNTSVNIIGFEGEKKLANHWSVAQMADEYASLLGLLQHQHAGHPPEELNRVIMCS 67
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 5 VDGNTHSVFSITGEGTFRWRSLTGTFQTEDELFSLHPLLL---GDAMREIKGIGVAS 61
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 68 VVPELTTFEERVKSKAPLVVAGIKSG-VKVRMDNPREVGADRIVNAARVLYPG 126
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 62 VVPQNTVIERFSQYFIHSPWKA--KNGCVKWNKPNSEVGADRVANVAVKEYGK 119
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 127 ACIIIVDMGTATFTDLSGGAYIGGAIAPGIATSAQAEKTSKPKIEIRPAK-VIGS 185
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 120 NGIIIDMTATVD-LVNGSYEGGAILPGFMMVHSIFRGCTAKPLPLVE-VKPADFVVGK 177
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 186 NTVSAMOSGGYFGYIGLVEELVRRIQTELG 215
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 178 DTEENIRLVNGSVYALEGIIGRIKEYG 207
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
RESULT 13
70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: F70165
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, son, D.; Peterson, J.; Kraljic, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70165
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <KLE>
A:Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g2688431
A:Experimental source: strain B31

Query Match 17.4%; Score 190; DB 2; Length 262;
Best Local Similarity 25.0%; Pred. No. 2e-08;
Matches 58; Conservative 45; Mismatches 87; Indels 42; Gaps 6;

Qy 1 MSEKLVA---VDIGNTSVNIIGFEGEKKLANHWSVAQMADEYASLLGLLQHQHAGHP 57
||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 1 MNKPLLSLLIIDIGNTSIAFALF-----DNQVNLFIKMKTNLMRY 42

Qy 58 EE-----LNRVIMCSVVPPLTTTFFEEVFKSYKAAAPLVVVGAGIKSGVKVRMDN 105
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 43 DEVISFFEEFNDFNVNKVFISVVPILNETFKNVIFSEFKKPLFGFDLNDYDLTF--N 99
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 106 PRE-----VGADRIVNAARVLYP-GACIIIVDMGTATFTDLSGGAYIGGAIAPGIAT 159
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 100 PYKSDKFLGSDVFANLVAJAENYSFENVLVVDLGTACTIFAVSRQDGLGILGINSGLPI 159
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 160 SAQAIKTSKPKIEIRPAKVIKSNVTSVAMOSGGYFGYIGLVEELVRRIQ 211
: : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 160 NFNSLLDNAYLIKKPEIPTPNLLERTTSGSVNGLFYQYKYLIEGVYDRIK 211
: : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
RESULT 14
A12292
hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: A12292
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriku, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12292
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-276 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075595.1; PID:g17133030; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3896

Query Match 14.6%; Score 160; DB 2; Length 276;
Best Local Similarity 24.7%; Pred. No. 6.6e-06;
Matches 55; Conservative 42; Mismatches 74; Indels 52; Gaps 9;

Qy 6 VAVDIGNTSVNIGFEGEKKLANHNL---GSVAORMADEYASLLGLLQHQHAGHPPEELN 61
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 15 LALEIGNSRHLWALFMGESLEFTWDTYLPESVIOQLNGETKLEVG-----SEERE 66
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 62 -----RVIMCSVVPPLTTTFFEEVFKSYKAAAPLVVVGAGIKSGVKVRMDN- 105
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 67 IFFTFPLPPAPCPPLPLFIASVVPQTVLWE---NLYNVRVITLD-----QIPLNNI 115
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 106 -PREVGADRIVNAARVLYPCACIIIVDMGTATFTDLSGGAYIGGAIAPGIATSAQAI 164
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 116 YP-TLIGIDRALALMGAGMSWGFVPLVIDAGTALTFTAADGGKMLVGGAILPGVGLQFASL 174
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Qy 165 AEKTSKPKIEIRPAKVIGS-----NTVSAMOSGGYFGYI 200
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
Db 175 GQQTGQLPQVEM----EAIKSLPPRFALNTTEAIOGSGVIYTLI 213
:||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : : ||:||||| : : : : :
RESULT 15
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitidis
N:Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C:Accession: H82031
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mungall, K.; Leather, S.; Moule, S.; Quail, M.A.; Rajandre, Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: AB1775; MUID:2022556; PMID:10761919
A:Accession: H82031
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-592 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: b1rA; NMA0357

C:Keywords: ligase

```
Query Match      13.9%; Score 152; DB 2; Length 592;
Best Local Similarity 24.3%; Pred. No. 7.3e-05;
Matches 53; Conservative 42; Mismatches 101; Indels 22; Gaps 5;

QY  2 SEKLVAVDIGNTSVNIGIPEGEKLLANWHLGSLVAQR-MADEYASLLGLLOHAGIHP 57
Db  | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
337 SERFLLDGNSRLKAWVNGTFAT---VGSAPYRDLSPGAEWAEEKVDGNV----- 386
QY  58 EELNRVIMCSVVPPLTTTTFEEVFKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNA 117
Db  | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
387 ---RIVGCAVCGEFKA-QVQEQRLARKIEWLPSSAQALGIRNHYRHPPEEHGSDRWENA 440
QY  118 AARVLYPGACIIIVDMGTATFTDTLSEGGAYGGATAPGIATSAQAIAEKTSKLPKIEII 177
Db  | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
441 LGSRRFSRNACVVVSCGTAVTDALTDGHHYLGTTIMPGFHLMKESLAVRTANLNRHAGK 500
QY  178 RPAKVIGSNTVSAMQSGIVFGYIGLVEELVRRIQTELG 215
Db  | | : | | : | | : | | : | | : | | : | | : | | : | | : | | :
501 R--YPPFTTGNVASCMDAVCGSVMMHGRUKETG 536
```

Search completed: June 24, 2003, 22:19:21

Job time : 10.0348 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 24.7896 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453A-57
Perfect score: 1093
Sequence: 1 MSEKLVAVDIGNTSVNIGIF.....IGLVEELVRRIQTELQKTR 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1093	100.0	219	AAU91176	Pantothenate kinas
2	558	51.1	255	AAU91154	Geobacter sulfurre
3	527	48.2	254	AAU91171	Pantothenate kinas
4	527	48.2	262	AAU91170	Pantothenate kinas
5	518	47.4	233	AAU91163	Pantothenate kinas
6	515	47.1	258	AAU01243	B. subtilis novel
7	515	47.1	258	AAU91149	Bacillus subtilis
8	513	46.9	250	AAU91150	Clostridium acetob
9	513	46.9	258	AAU91172	Pantothenate kinas
10	492.5	45.1	265	AAU91151	Streptomyces coeli

11	489	44.7	256	23	AAU91175	Pantothenate kinas
12	489	44.7	259	23	ABBA7661	Listeria monocytog
13	478	43.7	260	23	AAU91173	Pantothenate kinas
14	394	36.0	258	23	AAU91153	Rhodobacter capsul
15	371	33.9	272	22	RAG81225	Mycobacterium tube
16	371	33.9	272	23	AAU91152	Mycobacterium tube
17	348.5	31.9	212	23	AAU91177	Pantothenate kinas
18	333.5	30.5	262	23	AAU91155	Deinococcus radiop
19	310	28.4	273	23	AAU91157	Treponema pallidum
20	279.5	25.6	246	23	AAU91156	Thermotoga maritim
21	265.5	24.3	257	23	AAU91174	Pantothenate kinas
22	190	17.4	262	23	AAU91158	Borrelia burgdorfe
23	189	17.3	241	23	AAU91179	Pantothenate kinas
24	184	16.8	244	23	AAU91168	Pantothenate kinas
25	153	14.0	455	20	AAU91167	Neisseria gonorrhoe
26	153	14.0	455	21	AAU91166	Neisseria gonorrhoe
27	153	14.0	460	23	AAU91167	Neisseria gonorrhoe
28	153	14.0	592	20	AAU91167	Neisseria gonorrhoe
29	153	14.0	592	21	AAU91167	Neisseria gonorrhoe
30	152	13.9	455	21	AAU91167	Neisseria meningit
31	152	13.9	592	20	AAU91166	Neisseria meningit
32	152	13.9	592	21	AAU91166	Neisseria meningit
33	152	13.9	592	23	AAU91166	Neisseria meningit
34	150.5	13.8	389	21	AAU91166	Neisseria meningit
35	150	13.7	592	20	AAU91166	Neisseria meningit
36	150	13.7	592	21	AAU91166	Neisseria meningit
37	150	13.7	592	23	AAU91166	Neisseria meningit
38	144	13.2	257	23	AAU91160	Synechocystis pant
39	143.5	13.1	242	23	AAU91180	Pantothenate kinas
40	137.5	12.6	189	20	AAU91164	Neisseria meningit
41	136	12.4	248	23	AAU91164	Pantothenate kinas
42	134.5	12.3	249	23	AAU91182	Pantothenate kinas
43	134	12.3	249	23	AAU91178	Pantothenate kinas
44	130.5	11.9	267	23	AAU91162	Bordetella pertussis
45	130	11.9	229	23	AAU91159	Aquifex aeolicus p

ALIGNMENTS

RESULT 1
AAU91176
ID AAU91176 standard; Protein; 219 AA.

AC AAU91176;
XX
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #14.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Dehalococcoides ethenogenes.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDE; ABK54197.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein -
PS Claim 10; Page 106-107; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 219 AA;

Query Match 100.0%; Score 1093; DB 23; Length 219;
Best Local Similarity 100.0%; Pred. No. 7e-108;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MSELKLVAVDIGNTSVNIIGIFEGEKLKLANHILGSAORMADEYASLLGLLQHAGIHPPEEL 60
1 MSELKLVAVDIGNTSVNIIGIFEGEKLKLANHILGSAORMADEYASLLGLLQHAGIHPPEEL 60
61 NRVMGSAVPPPLTTTFFEEVFKSYFKAAPLVVGGAGIKSGVKVRMDNPREVGADRVNAAA 120
61 NRVMGSAVPPPLTTTFFEEVFKSYFKAAPLVVGGAGIKSGVKVRMDNPREVGADRVNAAA 120
121 RVLYPGACIIIVDMGTATFTLSEGGAYIGGATAPGIATSAQAIKTSKLPKIEIIRPA 180
121 RVLYPGACIIIVDMGTATFTLSEGGAYIGGATAPGIATSAQAIKTSKLPKIEIIRPA 180
181 KVIKGSNTVSAMQSGIYFGYIGLVEELVRRITQELGOKTR 219
181 KVIKGSNTVSAMQSGIYFGYIGLVEELVRRITQELGOKTR 219

RESULT 2
AAU91154
ID AAU91154 standard; Protein; 255 AA.
AC AAU91154;
XX
DT 05-JUN-2002 (first entry)
XX
DE Geobacter sulfurreducens pantothenate kinase Coax.
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.
Geobacter sulfurreducens.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
PS Claim 10; Page 72-73; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 255 AA;

Query Match 51.1%; Score 558; DB 23; Length 255;
Best Local Similarity 49.8%; Pred. No. 6.7e-51;
Matches 104; Conservative 47; Mismatches 58; Indels 0; Gaps 0;
5 LVAVDIGNTSVNIIGIFEGEKLKLANHILGSAORMADEYASLLGLLQHAGIHPPEELNRVI 64
2 LLVIDVGNINVLGIYDGERLVRDVRSTDKARTTDEYGLINELFRLAGLDQIRAVI 61
65 MCSVVPPLTTTFFEEVFKSYFKAAPLVVGGAGIKSGVKVRMDNPREVGADRVNAAAARVLY 124
62 ISSVVPPLTGVLRLSLGIFGMRPLVVGPIGTGMPIDNPREVGDRIVNAVAGYEKY 121
125 PGACIIIVDMGTATFTLSEGGAYIGGATAPGIATSAQAIKTSKLPKIEIIRPAKVI 184
122 RYSLIIVDGTATFTFVYVNRKGEYCGGATAPGLVISTEALFQASKLRPRVDIIRPSAII 181
185 SNTVSAMQSGIYFGYIGLVEELVRRITQTE 213
182 RNTVNSMQAGIYGYVGLVDEIVTRMAE 210

RESULT 3
AAU91171
ID AAU91171 standard; Protein; 254 AA.
AC AAU91171;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #9.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus halodurans.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54192.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
PS Claim 10; Page 100; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic

comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 254 AA;
Query Match 48.2%; Score 527; DB 23; Length 254;
Best Local Similarity 43.5%; Pred. No. 1.3e-47;
Matches 91; Conservative 57; Mismatches 61; Indels 0; Gaps 0;
QY 5 LVAVDIGNTSVNIIGFEGEKLLANHLGSAQRADEYASLLGLLQHAGHPPELNRI 64
2 ILVDVGNNTNVLGVYQDETLVHHRLATSKQTEDEYAMTVRSFLFHAGLQFQDIDGIV 61
65 MCSVVPPLTTTFEEVFKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNAARVLY 124
62 ISSVVPMPFSLQMKCKYHVTPIIIGPGIKTGLNLIKYNPKVEGADRIVNAARVLY 121
125 PGACIIVDMGTATFTDLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAKVG 184
122 GYPAIVVDGFTATTYCLINEKKQYAGGVIAPIGMISTEALYHRAKSLPRIETAKPKQVVG 181
185 SNTVSAMQSGIYFGYIGLVEELVRRIQTE 213
182 TMTDSMQSGIFYGVQSDGVVKRMAQ 210

RESULT 4
AAU91170
ID AAU91170 standard; Protein; 262 AA.
XX AC AAU91170;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #8.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus anthracis.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX DR N-PSDB; ABK54191.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
Claim 8; Page 98-99; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability

of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 262 AA;
Query Match 48.2%; Score 527; DB 23; Length 262;
Best Local Similarity 47.0%; Pred. No. 1.4e-47;
Matches 101; Conservative 46; Mismatches 68; Indels 0; Gaps 0;
QY 5 LVAVDIGNTSVNIIGFEGEKLLANHLGSAQRADEYASLLGLLQHAGHPPELNRI 64
2 IFVLDVGNNTNVLGVYQDETLVHHRLATSKQTEDEYAMTVRSFLFHAGLQFQDIDGIV 61
65 MCSVVPPLTTTFEEVFKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNAARVLY 124
62 VSSVVPPIIMFALERMCKEYFKIKPLVVGPGIKTGLNLIKYNPREVGADRIVNAARVLY 121
125 PGACIIVDMGTATFTDLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAKVG 184
122 GSPLIIVDFGFTATTYCYINEEKHYMGVITFGIMISAEALYSRAAKLPRIETKPSVVG 181
185 SNTVSAMQSGIYFGYIGLVEELVRRIQTELGQTR 219
182 KNTVSAMQSGILYGVQGVGIVRKMEAKQEPK 216

RESULT 5
AAU91163
ID AAU91163 standard; Protein; 233 AA.
XX AC AAU91163;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #1.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus subtilis.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX DR N-PSDB; ABK54169.

Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -
Disclosure; Page 81-82; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a

CC agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

XX Sequence 260 AA;

Query Match 43.7%; Score 478; DB 23; Length 260; Best Local Similarity 44.9%; Pred. No. 2.2e-42; Matches 96; Conservative 41; Mismatches 75; Indels 2; Gaps 1;

QY 5 LVAVDIGNTSVNTGIFEGEKLKLANHLGWSVAORMADEYASLLGLLQHAGIHPEELNRVI 64

Db 2 LLAIEQNTNTMFAIHGASWVAQMSRSTADEYVWLSQLLSMQGLGFRDAVI 61

QY 65 MCSVVPPLTTTTEEFVKSKAAPLVVGGAGIKSGVKVRMDNPREVCADRIVNAAARVLY 124

Db 62 ISSVVPQSTFNLNLSRRYFNVPLVIGENAKIGIDVRIEKPSEAGADRLVNAIGAAMY 121

QY 125 PGACIIVDMGTATFTDLSEGGAYIGGATAPGATSAQAIAEKTSLPKIEIIRPA--KV 182

Db 122 PGLVVIDSGTATFDIADGAFEGGIAPGILNSQALHEAAKLPRIAIOPAGNRI 181

QY 183 IGSNTVSAMSGIYFGYIGLVEELVRRIOTELQ 216

Db 182 VGTDTVSAMSGVFWGYISLIEGLVARIKAE 215

RESULT 14

AAU91153

ID AAU91153 standard; Protein; 258 AA.

XX AC AAU91153;

XX 05-JUN-2002 (first entry)

XX Rhodobacter capsulatus pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Rhodobacter capsulatus.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises

XX contacting composition comprising pantothenate kinase (Coax) protein

XX with test compound and identifying inhibitor of the Coax protein -

XX Claim 10; Page 71-72; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

XX comprising contacting an assay composition comprising a pantothenate

XX kinase (Coax) protein with a test compound, and determining the ability

XX of the test compound to inhibit the activity of the Coax protein, an

CC described in the invention.

XX Sequence 258 AA;

Query Match 36.0%; Score 394; DB 23; Length 258;

Best Local Similarity 40.0%; Pred. No. 1.9e-33;

Matches 84; Conservative 39; Mismatches 85; Indels 2; Gaps 2;

QY 5 LVAVDIGNTSVNTGIFEGEKLKLANHLGWSVAORMADEYASLLGLLQHAGIHPEELNRVI 64

Db 2 LLDICDGNNTVTFVSDGDFATWRIATDHRRTADEYFWLNTLMQLKGLQ-GRISEAI 60

QY 65 MCSVVPPLTTTTEEFVKSKAAPLVVGGAGIKSGVKVRMDNPREVCADRIVNAAARVLY 123

Db 61 IISTAPRVVFNLRVLCNRYFDCRPVYVVGKPGCELPVAPRVDPTTTPVDPRLVNTVAGYDR 120

QY 124 YGACIIVDMGTATFTDLSEGGAYIGGATAPGATSAQAIAEKTSLPKIEIIRPAKV 183

Db 121 HGGDLIVVDGTFATFDVVPADGAYTGGVIAPGVNLSLEALHMAAALPHVDVTKPQGYI 180

QY 184 GSNVTSAMSGIYFGYIGLVEELVRRIOE 213

Db 181 GINTVACIOGVIWGYIGLVEGIVRQIRME 210

RESULT 15

AAG81225

ID AAG81225 standard; Protein; 272 AA.

XX AC AAG81225;

XX 04-SEP-2001 (first entry)

XX Mycobacterium tuberculosis potential drug target protein SEQ ID 276.

XX Drug target; growth; organism viability; characterisation.

XX Mycobacterium tuberculosis.

XX WO200135317-A1.

XX 17-MAY-2001.

XX 13-NOV-2000; 2000WO-US31152.

XX 12-NOV-1999; 99US-0165086.

XX 12-NOV-1999; 99US-0165124.

XX 01-FEB-2000; 2000US-0179531.

XX (REGC) UNIV CALIFORNIA.

XX Eisenberg D, Rotstein SH, Marcotte EM;

XX WPI; 2001-329193/34.

XX N-PSDB; AAH52076.

XX Identifying nucleotide or polypeptide sequence for use as drug target,

XX involves providing algorithm that analyzes a functional relationship

XX between nucleotide or polypeptide sequences, and comparing the

XX sequences -

XX Disclosure; Page 188; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or

XX polypeptide sequence that may be a drug target, or essential for growth

CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium

CC tuberculosis proteins which are potential drug targets. The DNA and

CC protein sequences are used to illustrate the method of the invention. The

CC method involves providing an unknown nucleotide or polypeptide sequences,

CC and comparing it to a number of sequences along with at least one

CC algorithm capable of analysing a functional relationship between

CC nucleotide and polypeptide sequences. The method is useful for

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 28.9778 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453A-55
Perfect score: 1305
Sequence: 1 MLLVDFVGNWMLGIYKGD.....YVDGFLTEGLRIIYERKNOE 256

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1305	100.0	256	AAU91175	Pantothenate kinas
2	817.5	62.6	258	AAU91172	Pantothenate kinas
3	814.5	62.4	254	AAU91171	Pantothenate kinas
4	802.5	61.5	262	AAU91170	Pantothenate kinas
5	794.5	60.9	258	AAU01243	B. subtilis novel
6	794.5	60.9	258	AAU91149	Bacillus subtilis
7	785.5	60.2	255	AAU91154	Geobacter sulfurre
8	779.5	59.7	259	ABR47661	Listeria monocytog
9	695	53.3	250	AAU91150	Clostridium acetob
10	672	51.5	233	AAU91163	Pantothenate kinas

11	643.5	49.3	265	23	AAU91151	Streptomyces coeli
12	523.5	40.1	260	23	AAU91173	Pantothenate kinas
13	513	39.3	272	23	AAU91152	Mycobacterium tube
14	509	39.0	272	22	AAU91153	Mycobacterium tube
15	496.5	38.0	258	23	AAU91153	Rhodobacter capsul
16	489	37.5	219	23	AAU91176	Pantothenate kinas
17	377	28.9	262	23	AAU91155	Deinococcus radiop
18	354	27.1	273	23	AAU91157	Treponema pallidum
19	321	24.6	246	23	AAU91156	Thermotoga maritim
20	315.5	24.2	212	23	AAU91177	Pantothenate kinas
21	282	21.6	262	23	AAU91158	Borrelia burgdorfe
22	254.5	19.5	257	23	AAU91174	Pantothenate kinas
23	201.5	15.4	244	23	AAU91168	Pantothenate kinas
24	145	11.1	241	23	AAU91179	Pantothenate kinas
25	144.5	11.1	257	23	AAU91160	Synechocystis pant
26	140.5	10.8	389	21	AAU914909	Neisseria meningit
27	140.5	10.8	455	21	AAU914910	Neisseria meningit
28	140.5	10.8	592	20	AAU9138615	Neisseria meningit
29	140.5	10.8	592	21	AAU9138616	Neisseria meningit
30	140.5	10.8	592	21	AAU914912	Neisseria meningit
31	140.5	10.8	592	21	AAU914913	Neisseria meningit
32	140.5	10.8	592	23	AAU91166	Pantothenate kinas
33	140.5	10.8	592	23	AAU91169	Pantothenate kinas
34	138.5	10.6	455	20	AAU9138617	Neisseria gonorrhoe
35	138.5	10.6	455	21	AAU9138618	Neisseria gonorrhoe
36	138.5	10.6	460	23	AAU91167	Pantothenate kinas
37	138.5	10.6	592	20	AAU9138618	Neisseria gonorrhoe
38	138.5	10.6	592	21	AAU914911	Neisseria gonorrhoe
39	133	10.2	229	23	AAU91159	Aquifex aeolicus p
40	131	10.0	249	23	AAU91178	Pantothenate kinas
41	127.5	9.8	267	23	AAU91162	Bordetella pertussis
42	127	9.7	249	23	AAU91182	Pantothenate kinas
43	120	9.2	242	23	AAU91180	Pantothenate kinas
44	119.5	9.2	248	23	AAU91164	Pantothenate kinas
45	110.5	8.5	189	20	AAU9138614	Neisseria meningit

ALIGNMENTS

RESULT 1
AAU91175
ID AAU91175 standard; Protein; 256 AA.
XX AAU91175;
XX
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #13.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Clostridium difficile.
OS
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54196.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein
 XX Claim 6; Page 105; 128pp; English.
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 XX described in the invention.
 SQ Sequence 256 AA;
 Query Match 100.0%; Score 1305; DB 23; Length 256;
 Best Local Similarity 100.0%; Pred. No. 8.5e-122;
 Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MLLVFDVGNMVLGIYKGDKNVYRIKTDREKTSDEYGLISNLFDDYDNVNISDIDDV 60
 1 MLLVFDVGNMVLGIYKGDKNVYRIKTDREKTSDEYGLISNLFDDYDNVNISDIDDV 60
 61 IISVVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQGVADRVNAVAGIEK 120
 61 IISVVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQGVADRVNAVAGIEK 120
 121 YGAPSLVDFGTATTCATSEKGEYLGGTAPGKIISSEALFOSAKLPRVELAKPGMTI 180
 121 YGAPSLVDFGTATTCATSEKGEYLGGTAPGKIISSEALFOSAKLPRVELAKPGMTI 180
 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
 241 FLTLEGLRIIYEKNQE 256
 241 FLTLEGLRIIYEKNQE 256
 RESULT 2
 AAU91172
 ID AAU91172 standard; Protein; 258 AA.
 XX AAU91172;
 AC AAU91172;
 DT 05-JUN-2002 (first entry)
 Pantothenate kinase (Coax) #10.
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 OS Bacillus stearothermophilus.
 XX WO200216601-A2.
 XX 28-FEB-2002.
 XX 24-AUG-2001; 2001WO-US26531.
 XX 24-AUG-2000; 2000US-227860P.
 XX 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.
 XX N-PSDB; ABK54193.

PT Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein
 XX Claim 10; Page 101-102; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 XX described in the invention.
 SQ Sequence 258 AA;
 Query Match 62.6%; Score 817.5; DB 23; Length 258;
 Best Local Similarity 62.9%; Pred. No. 3.5e-73;
 Matches 161; Conservative 40; Mismatches 54; Indels 1; Gaps 1;
 1 MLLVFDVGNMVLGIYKGDKNVYRIKTDREKTSDEYGLISNLFDDYDNVNISDIDDV 60
 1 MIFVLDVGNMVLGIYKGDKNVYRIKTDREKTSDEYGLISNLFDDYDNVNISDIDDV 60
 61 IISVVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQGVADRVNAVAGIEK 120
 61 IISVVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQGVADRVNAVAGIEK 120
 121 YGAPSLVDFGTATTCATSEKGEYLGGTAPGKIISSEALFOSAKLPRVELAKPGMTI 180
 121 YGAPSLVDFGTATTCATSEKGEYLGGTAPGKIISSEALFOSAKLPRVELAKPGMTI 180
 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
 241 FLTLEGLRIIYEKNQE 256
 240 FLTLEGLRIIYEKNQE 255
 RESULT 3
 AAU91171
 ID AAU91171 standard; Protein; 254 AA.
 XX AAU91171;
 AC AAU91171;
 DT 05-JUN-2002 (first entry)
 Pantothenate kinase (Coax) #9.
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 OS Bacillus halodurans.
 XX WO200216601-A2.
 XX 28-FEB-2002.
 XX 24-AUG-2001; 2001WO-US26531.
 XX 24-AUG-2000; 2000US-227860P.
 XX 20-MAR-2001; 2001US-0813453.
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.

DR N-PSDB; ABK54192.
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
XX Claim 10; Page 100; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 254 AA;
Query Match 62.4%; Score 814.5; DB 23; Length 254;
Best Local Similarity 60.2%; Pred. No. 6.9e-73;
Matches 153; Conservative 45; Mismatches 55; Indels 1; Gaps 1;
Qy 1 MLLVDFGNTNMVLGIYKGDKNVYWRKTDREKTSDEYGLISNLFYDYNVNSIDDDV 60
Db 1 MILVIDGNTNVLGVYQDETIVHWRLATSRQKTEDEYAMTVRSDFDHAGLQFQDIDGI 60
Qy 61 IISVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQVQADRIYNVAVAGIEK 120
Db 61 VISSVPPMFLSOMCKYFHVTPMTIGPGIKTGLNPKYDNPKEVGADRIYNVAVATEL 120
Qy 121 YGAPSIIVDFGTATTCFAISEKGEYLGVTIAPGKISSEALFQSAKLPRVELAKPGWTI 180
Db 121 YGPAIVVDFGTATTCFAISEKGEYLGVTIAPGKISSEALFQSAKLPRVEIAKPKQV 180
Qy 181 CKSTVSAQSGIYGYVGLVDKIISIMKELNCDVVKVIATGGLAKLITASETKSIDYVDG 240
Db 181 GTNFDMSQSGIFGYVQVGVVGRMKQAQE-SEPKVIATGGLAKLITASETKSIDVIDS 239
Qy 241 FLLEGLRIIYEKN 254
Db 240 FLTLKGLQLYKKN 253
RESULT 4
AAU01170 standard; Protein; 262 AA.
XX AAU01170;
XX AC AAU01170;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #8.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus anthracis.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.
DR N-PSDB; ABK54191.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
XX Claim 8; Page 98-99; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 262 AA;
Query Match 61.5%; Score 802.5; DB 23; Length 262;
Best Local Similarity 59.1%; Pred. No. 1.1e-71;
Matches 150; Conservative 49; Mismatches 54; Indels 1; Gaps 1;
Qy 1 MLLVDFGNTNMVLGIYKGDKNVYWRKTDREKTSDEYGLISNLFYDYNVNSIDDDV 60
Db 1 MIFVLDVGNNAVJGVFEEGELRQHRWETDRHKTEDEYGLMLVKQLLEHGLSFDVKG 60
Qy 61 IISVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQVQADRIYNVAVAGIEK 120
Db 61 IVSSVPPMFLSOMCKYFHVTPMTIGPGIKTGLNPKYDNPKEVGADRIYNVAVAGIHL 120
Qy 121 YGAPSIIVDFGTATTCFAISEKGEYLGVTIAPGKISSEALFQSAKLPRVELAKPGWTI 180
Db 121 YGSPILVDFGTATTCFAISEKGEYLGVTIAPGKISSEALFQSAKLPRVEIAKPKQV 180
Qy 181 CKSTVSAQSGIYGYVGLVDKIISIMKELNCDVVKVIATGGLAKLITASETKSIDYVDG 240
Db 181 KNTVSAQSGIYGYVGLVDKIISIMKELNCDVVKVIATGGLAKLITASETKSIDYVDV 239
Qy 241 FLLEGLRIIYEKN 254
Db 240 FLTLKGLQLYKKN 253
RESULT 5
AAU01243
ID AAU01243 standard; Protein; 258 AA.
XX AC AAU01243;
XX DT 18-JUL-2001 (first entry)
XX DE B. subtilis novel pantothenate kinase encoded by the gene coax.
XX KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;
XX KW nutritional supplement; panto-compound; pantoate.
XX OS Bacillus subtilis.
XX PN WO200121772-A2.
XX PD 29-MAR-2001.
XX PF 21-SEP-2000; 2000WO-US25993.
XX PR 21-SEP-1999; 99US-0400494.
XX PR 07-JUN-2000; 2000US-0210072.
XX PR 28-JUL-2000; 2000US-0221836.
XX PR 24-AUG-2000; 2000US-0227860.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS.
XX PI Yocum RR, Patterson TA, Hermann T, Pero JG;
XX PR WPI; 2001-218644/22.
XX DR N-PSDB; AAS00984.
XX XX
XX PT New recombinant microorganism which overexpress a Bacillus subtilis
XX PT pantothenate biosynthetic enzyme, useful for the high yield production
XX PT of panto-compounds such as pantothenate and pantoate -
XX PS Example 14; Fig 23; 292pp; English.
XX CC The sequence represents a novel B. subtilis pantothenate kinase (encoded
XX CC by gene cox), an enzyme of the pantothenate biosynthetic pathway.
XX CC Pantothenate, also known as vitamin B5, is used as a nutritional
XX CC supplement in mammals and humans. The invention concerns methods of
XX CC producing recombinant microorganisms overexpressing at least one Bacillus
XX CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods
XX CC of producing them are useful for producing a panto-compound such as
XX CC pantothenate or pantoate, which is a nutritional requirement for
XX CC livestock and humans. The methods are also useful for the identification
XX CC of pantothenate kinase modulators. Panto-compounds are produced at a
XX CC significantly higher yield than prior art methods and can be produced
XX CC independent of the need to feed precursors which decreases expense.
XX XX
XX SQ Sequence 258 AA;

Query Match 60.9%; Score 794.5; DB 22; Length 258;
Best Local Similarity 60.8%; Pred. No. 7e-71;
Matches 155; Conservative 40; Mismatches 59; Indels 1; Gaps 1;
QY 1 MLLVDVGNNTVMVLGYKDGKLVNWRKTDREKTSDEYGLISNLFYDYNVISDIDV 60
DB 1 LLLVIDVGNNTVMVLGYHDKLEYHWRKTSDEFGMLRSLFDSGLMFEQIDGI 60
QY 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQVGDRIYVAVAGIEK 120
DB 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQVGDRIYVAVAAIHL 120
QY 121 YGAPSLVDFGTTATTCFAISERKGLGTTAPCIKISSALFQSAKLPRVELAKPGMTI 180
DB 121 YGNPLIVDFGTTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
QY 181 CRSTVSAMQSGIYGVGLVDKIISTMKKELNCDVVKVIATGGLAKLIASETKSIDYVDG 240
DB 181 GRNTVSAMQSGILFGVGVQGVGIVKRMKQWAK-QDLKVIATGGLAPLIANESDCIDIVDP 239
241 FLTLEGLRIIYEKNO 255
240 FLTLEGLRIIYERNR 254

RESULT 6
AAU91149
ID AAU91149 standard; Protein; 258 AA.
XX AC AAU91149;
XX DT 05-JUN-2002 (first entry)
XX DE Bacillus subtilis pantothenate kinase Coax.
XX DE Geobacter sulfurreducens pantothenate kinase Coax.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus subtilis.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX XX

PF 24-AUG-2001; 2001WO-US26531.
XX XX
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX XX
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX XX
XX PI Yocum RR, Patterson TA;
XX XX
XX DR WPI; 2002-269358/31.
XX DR N-PSDB; ABR54168.
XX XX
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -
XX PS Claim 10; Page 67-68; 128pp; English.
XX XX
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX XX
XX SQ Sequence 258 AA;

Query Match 60.9%; Score 794.5; DB 23; Length 258;
Best Local Similarity 60.8%; Pred. No. 7e-71;
Matches 155; Conservative 40; Mismatches 59; Indels 1; Gaps 1;
QY 1 MLLVDVGNNTVMVLGYKDGKLVNWRKTDREKTSDEYGLISNLFYDYNVISDIDV 60
DB 1 LLLVIDVGNNTVMVLGYHDKLEYHWRKTSDEFGMLRSLFDSGLMFEQIDGI 60
QY 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQVGDRIYVAVAGIEK 120
DB 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNPKYDNPQVGDRIYVAVAAIHL 120
QY 121 YGAPSLVDFGTTATTCFAISERKGLGTTAPCIKISSALFQSAKLPRVELAKPGMTI 180
DB 121 YGNPLIVDFGTTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
QY 181 CRSTVSAMQSGIYGVGLVDKIISTMKKELNCDVVKVIATGGLAKLIASETKSIDYVDG 240
DB 181 GRNTVSAMQSGILFGVGVQGVGIVKRMKQWAK-QDLKVIATGGLAPLIANESDCIDIVDP 239
241 FLTLEGLRIIYEKNO 255
240 FLTLEGLRIIYERNR 254

RESULT 7
AAU91154
ID AAU91154 standard; Protein; 255 AA.
XX AC AAU91154;
XX DT 05-JUN-2002 (first entry)
XX DE Geobacter sulfurreducens pantothenate kinase Coax.
XX DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Geobacter sulfurreducens.
XX PN WO200216601-A2.
XX XX

PD 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
PR
XX 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
FA
XX Yocum RR, Patterson TA;
XX
XX WPI; 2002-269358/31.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein
XX
XX Claim 10; Page 72-73; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX
XX Sequence 255 AA;
XX
XX Query Match 60.2%; Score 785.5; DB 23; Length 255;
XX Best Local Similarity 60.9%; Pred. No. 5.4e-70;
XX Matches 156; Conservative 44; Mismatches 55; Indels 1; Gaps 1;
XX
Qy 1 MLLVFDVGNNTMVLGIYKDKLVNWRKTDREKTSDEYGLISNLFYDNNVNSIDDDV 60
Db 1 MLLVFDVGNNTMVLGIYKDKLVNWRKTDREKTSDEYGLISNLFYDNNVNSIDDDV 60
Qy 61 IISWVPMVHSLNFICIKYCKKQPLIVGPGIKTGLNPKYDNPQVGNADRVNAVAGIEK 120
Db 61 IISWVPMVHSLNFICIKYCKKQPLIVGPGIKTGLNPKYDNPQVGNADRVNAVAGIEK 120
Qy 121 YGAPSVLVDFGTATTCFCAISEKGEYLGTTAPGKIKISSEALFQSKLPRVELAKPGMTI 180
Db 121 YGAPSVLVDFGTATTCFCAISEKGEYLGTTAPGKIKISSEALFQSKLPRVELAKPGMTI 180
Qy 181 CKSTVSMQSGIYGYVGLVDKIISIMKKNLDDVVKVIATGGLAKLIASFTKSIDYVDG 240
Db 181 CKSTVSMQSGIYGYVGLVDKIISIMKKNLDDVVKVIATGGLAKLIASFTKSIDYVDG 240
Qy 241 FLTLEGLRIIYKNOE 255
Db 240 YLTLEGLRIIYERNKE 255
RESULT 8
ID ABB47661
XX ABB47661 standard; Protein; 259 AA.
XX
XX ABB47661;
XX
XX 05-FEB-2002 (first entry)
XX
XX Listeria monocytogenes protein #365.
DE
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX
XX Listeria monocytogenes.
OS
XX WO200177335-A2.
PN

XX 18-OCT-2001.
PD
XX 11-APR-2001; 2001WO-FR01118.
PF
XX 11-APR-2000; 2000FR-0004629.
PR
XX (INSP) INST PASTEUR.
XX
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Medjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Anand A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose N, Voss H;
XX
XX WPI; 2002-010914/01.
DR
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides
XX
XX Claim 6; SEQ ID No 366; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccines compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 259 AA;
XX
XX Query Match 59.7%; Score 779.5; DB 23; Length 259;
XX Best Local Similarity 56.9%; Pred. No. 2.2e-69;
XX Matches 145; Conservative 50; Mismatches 59; Indels 1; Gaps 1;
XX
Qy 1 MLLVFDVGNNTMVLGIYKDKLVNWRKTDREKTSDEYGLISNLFYDNNVNSIDDDV 60
Db 1 MLLVFDVGNNTMVLGIYKDKLVNWRKTDREKTSDEYGLISNLFYDNNVNSIDDDV 60
Qy 61 IISWVPMVHSLNFICIKYCKKQPLIVGPGIKTGLNPKYDNPQVGNADRVNAVAGIEK 120
Db 61 IISWVPMVHSLNFICIKYCKKQPLIVGPGIKTGLNPKYDNPQVGNADRVNAVAGIEK 120
Qy 121 YGAPSVLVDFGTATTCFCAISEKGEYLGTTAPGKIKISSEALFQSKLPRVELAKPGMTI 180
Db 121 YGAPSVLVDFGTATTCFCAISEKGEYLGTTAPGKIKISSEALFQSKLPRVELAKPGMTI 180
Qy 181 CKSTVSMQSGIYGYVGLVDKIISIMKKNLDDVVKVIATGGLAKLIASFTKSIDYVDG 240
Db 181 CKSTVSMQSGIYGYVGLVDKIISIMKKNLDDVVKVIATGGLAKLIASFTKSIDYVDG 240
Qy 241 FLTLEGLRIIYKNOE 255
Db 240 FLTLEGLRIIYERNKE 255
RESULT 9

AAU91150
 ID AAU91150 standard; Protein; 250 AA.
 AC AAU91150;
 XX
 DT 05-JUN-2002 (first entry)
 AC
 XX
 DE Clostridium acetobutylicum pantothenate kinase Coax.
 DE
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 KW
 XX Clostridium acetobutylicum.
 OS
 XX WO200216601-A2.
 PN
 XX 28-FEB-2002.
 PD
 XX 24-AUG-2001; 2001WO-US26531.
 XX
 PF 24-AUG-2000; 2000US-227860P.
 XX
 PR 20-MAR-2001; 2001US-0813453.
 XX
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX
 PI Yocum RR, Patterson TA;
 XX
 DR WPI; 2002-269358/31.
 XX
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 PS Claim 10; Page 68-69; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of the Coax protein, an
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 250 AA;
 Query Match 53.3%; Score 695; DB 23; Length 250;
 Best Local Similarity 59.0%; Pred. No. 5.6e-61;
 Matches 138; Conservative 37; Mismatches 51; Indels 8; Gaps 2;
 QY 1 MLVFDVGNNTNVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFYDYNVISDIDDV 60
 DB 18 VILVDVGNNTNVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFYDYNVISDIDDV 77
 QY 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNRYDNPQVGDRIYNAVAGIEK 120
 DB 78 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNRYDNPQVGDRIYNAVAGIEK 137
 QY 121 YGAPSLVDFGRTATTCATSEKGEYLGTTAPGKIISSEALFQSAKSLPRVELAKPGMTI 180
 DB 138 YKRSLLIIDFGTATTCATSEKGEYLGTTAPGKIISSEALFQSAKSLPRVELAKPGMTI 197
 QY 181 CKSTVSAMQSGIIVGVGLVDKLIISIMKELNCDDVK-----VATGGGLAKLI 228
 DB 198 CANTISSIQSGIIVRYLQVYLFKLEK--NLPDGRTRTSLVLTGGLAKLI 249
 RESULT 10
 AAU91163
 ID AAU91163 standard; Protein; 233 AA.
 AC AAU91163;
 XX
 DT 05-JUN-2002 (first entry)
 AC
 XX
 DE Pantothenate kinase (Coax) #1.
 DE
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 KW
 XX Bacillus subtilis.
 OS
 XX WO200216601-A2.
 PN
 XX 28-FEB-2002.
 PD
 XX 24-AUG-2001; 2001WO-US26531.
 XX
 PF 24-AUG-2000; 2000US-227860P.
 XX
 PR 20-MAR-2001; 2001US-0813453.
 XX
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX
 PI Yocum RR, Patterson TA;
 XX
 DR WPI; 2002-269358/31.
 XX
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 PS Disclosure; Page 81-82; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of the Coax protein, an
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 233 AA;
 Query Match 51.5%; Score 672; DB 23; Length 233;
 Best Local Similarity 60.3%; Pred. No. 1e-58;
 Matches 129; Conservative 32; Mismatches 53; Indels 0; Gaps 0;
 QY 1 MLVFDVGNNTNVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFYDYNVISDIDDV 60
 DB 1 MLVFDVGNNTNVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFYDYNVISDIDDV 60
 QY 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNRYDNPQVGDRIYNAVAGIEK 120
 DB 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNRYDNPQVGDRIYNAVAGIEK 120
 QY 121 YGAPSLVDFGRTATTCATSEKGEYLGTTAPGKIISSEALFQSAKSLPRVELAKPGMTI 180
 DB 121 YGAPSLVDFGRTATTCATSEKGEYLGTTAPGKIISSEALFQSAKSLPRVELAKPGMTI 180
 QY 181 CKSTVSAMQSGIIVGVGLVDKLIISIMKELNCDDVK-----VATGGGLAKLI 214
 DB 181 CKSTVSAMQSGIIVGVGLVDKLIISIMKELNCDDVK-----VATGGGLAKLI 214
 RESULT 11
 AAU91151
 ID AAU91151 standard; Protein; 265 AA.
 XX
 AC AAU91151;
 XX

AC AAU91163;
 XX
 DT 05-JUN-2002 (first entry)
 AC
 XX
 DE Pantothenate kinase (Coax) #1.
 DE
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 KW
 XX Bacillus subtilis.
 OS
 XX WO200216601-A2.
 PN
 XX 28-FEB-2002.
 PD
 XX 24-AUG-2001; 2001WO-US26531.
 XX
 PF 24-AUG-2000; 2000US-227860P.
 XX
 PR 20-MAR-2001; 2001US-0813453.
 XX
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX
 PI Yocum RR, Patterson TA;
 XX
 DR WPI; 2002-269358/31.
 XX
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 PS Disclosure; Page 81-82; 128pp; English.
 XX
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of the Coax protein, an
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX
 SQ Sequence 233 AA;
 Query Match 51.5%; Score 672; DB 23; Length 233;
 Best Local Similarity 60.3%; Pred. No. 1e-58;
 Matches 129; Conservative 32; Mismatches 53; Indels 0; Gaps 0;
 QY 1 MLVFDVGNNTNVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFYDYNVISDIDDV 60
 DB 1 MLVFDVGNNTNVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFYDYNVISDIDDV 60
 QY 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNRYDNPQVGDRIYNAVAGIEK 120
 DB 61 IISVVVPMVHSLNFCIKYCKKQPLIVGPGIKTGLNRYDNPQVGDRIYNAVAGIEK 120
 QY 121 YGAPSLVDFGRTATTCATSEKGEYLGTTAPGKIISSEALFQSAKSLPRVELAKPGMTI 180
 DB 121 YGAPSLVDFGRTATTCATSEKGEYLGTTAPGKIISSEALFQSAKSLPRVELAKPGMTI 180
 QY 181 CKSTVSAMQSGIIVGVGLVDKLIISIMKELNCDDVK-----VATGGGLAKLI 214
 DB 181 CKSTVSAMQSGIIVGVGLVDKLIISIMKELNCDDVK-----VATGGGLAKLI 214
 RESULT 11
 AAU91151
 ID AAU91151 standard; Protein; 265 AA.
 XX
 AC AAU91151;
 XX

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.5612 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453a-55

Perfect score: 1305

Sequence: 1 MLLVFDVGNVNLGIYKGD.....YVDGFLTEGLRIIYEKNQE 256

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	826.5	63.3	273	2 E97293	probable transcript
2	814.5	62.4	254	2 F83660	hypothetical prote
3	779.5	59.7	259	2 AF1102	conserved hypothet
4	777.5	59.6	259	2 AF1464	conserved hypothet
5	672	51.5	233	2 S66100	conserved hypothet
6	643.5	49.3	265	2 T36391	hypothetical prote
7	529	40.5	274	2 H86937	conserved hypothet
8	523.5	40.1	261	2 B87489	transcription acti
9	513	39.3	272	2 A70955	hypothetical prote
10	377	28.9	262	2 E75516	conserved hypothet
11	354	27.1	273	2 D71326	conserved hypothet
12	321	24.6	246	2 D72320	conserved hypothet
13	282	21.6	262	2 F70165	conserved hypothet
14	156	12.0	276	2 A12292	hypothetical prote
15	144.5	11.1	257	2 S75559	hypothetical prote
16	140.5	10.8	592	2 H81009	Bira protein/Bvg a
17	140.5	10.8	592	2 H82031	probable biotin-[a
18	133	10.2	229	2 E70465	hypothetical prote
19	127.5	9.8	267	2 I40327	baf protein - Bord
20	120	9.2	242	2 A82637	conserved hypothet
21	119.5	9.2	248	2 H83111	hypothetical prote
22	115	8.8	223	2 G71887	hypothetical prote
23	114.5	8.8	224	2 A99571	conserved hypothet
24	105	8.0	252	2 T28315	ORF MSV154 probabl
25	102.5	7.9	1159	1 H64089	DNA-directed DNA p
26	101	7.7	223	2 F64627	hypothetical prote
27	98	7.5	597	2 B69251	probable electron
28	96	7.4	660	2 E90576	conserved hypothet
29	95.5	7.3	513	2 T09739	aspartic endopepti

30	95	7.3	1161	2 A84958	DNA-directed DNA p
31	92.5	7.1	270	2 G97299	pyrrolidine-5-carbox
32	92.5	7.1	381	2 T03289	formaldehyde dehyd
33	92.5	7.1	1036	2 F81719	isoleucyl-tRNA syn
34	92	7.0	566	2 F97224	ion-like ATP-depen
35	91.5	7.0	381	2 T04164	formaldehyde dehyd
36	91	7.0	938	2 AB1530	transcriptional regu
37	90.5	6.9	735	2 S74209	multifunctional be
38	90	6.9	938	2 A11172	transcriptional regu
39	89	6.8	393	2 A64036	hypothetical prote
40	88.5	6.8	279	2 T32196	hypothetical prote
41	88.5	6.8	990	2 T02309	probable transcript
42	88	6.7	553	1 VGN201	cell fusion glycop
43	87.5	6.7	247	2 S73535	hypothetical prote
44	87.5	6.7	1160	2 A10530	DNA polymerase III
45	87.5	6.7	1164	2 G82100	DNA polymerase III

ALIGNMENTS

RESULT 1

E97293

probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostr

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: E97293

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: E97293

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 <R>

A:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3200

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match 63.3%; Score 826.5; DB 2; Length 273;

Best Local Similarity 62.0%; Pred. No. 1.1e-59;

Matches 160; Conservative 42; Mismatches 53; Indels 3; Gaps 1;

QY 1 MLLVFDVGNVNLGIYKDKLVNWRKTDREKTSDEYGLISNLFYDNNISDIDDV 60

Db	12	VILVLDVGNVNLGIYNDTKLTAEWRLSTDVLSADEYGIQVMNLFQDQKLDPTLVEGV 71
QY	61	IISSVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGNIKYDNPKNQVQADRVNAVAGIEK 120
Db	72	IISSVVPNMVHSLNFCIKYCKKQPLIVGPGIKTGNIKYDNPKNQVQADRVNAVAGIEK 131
QY	121	YGAPSLVDGFTATTCFAISGEYLGTTAFGKISSEALFQSAKLPVELAKPGMT 180
Db	132	YKRSLLIIDFGTATTCFAISGEYLGTTAFGKISSEALFQSAKLPVELAKPGMT 191
QY	181	CKSTVSAMSGIIYGVGLVDKIIISTMKRELNCDDVK---VIATGGLAKLIASETSIDY 237
Db	192	CKNTISSIGVIYGVGLVDKIIISTMKRELNCDDVK---VIATGGLAKLIASETSIDY 251
QY	238	VDGFLTEGLRIIYEKNQ 255
Db	252	INPFLTEGLRIIYEKNR 269

RESULT 2

F83660

hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: F83660

A:Reference number: S65967; MUID:96051385; PMID:7584024
A:Accession: S66100
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <OGA>
A:Cross-references: EMBL:D26185; NID:9467326; PIDN:BA05305.1; PID:dl005847; PID:9467459
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardinols
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanton
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A95580; MUID:98044033; PMID:9384377
A:Accession: E69740
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <KUN>
A:Cross-references: GB:299104; GB:AL009126; NID:92632267; PIDN:CAB11846.1; PID:el182003;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yacB
A:Start codon: TTG
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 51.5%; Score 672; DB 2; Length 233;
Best Local Similarity 60.3%; Pred. No. 3.2e-47;
Matches 129; Conservative 32; Mismatches 53; Indels 0; Gaps 0;
RESULT 6
Query Match 51.5%; Score 672; DB 2; Length 233;
Best Local Similarity 60.3%; Pred. No. 3.2e-47;
Matches 129; Conservative 32; Mismatches 53; Indels 0; Gaps 0;
QY 1 MLLVFDVGNNTNVLGIYKDKLVNRYIKTDRKTSDEYGLISNLFYDYNVNSIDDDV 60
DB 1 MLLVFDVGNNTNVLGIYKDKLVNRYIKTDRKTSDEYGLISNLFYDYNVNSIDDDV 60
QY 61 IISVVVPMFALERMCTKYFHEIQVPGMKTGNIKYDNPKQVADRIYNAVAGIEK 120
DB 61 IISVVVPMFALERMCTKYFHEIQVPGMKTGNIKYDNPKQVADRIYNAVAGIEK 120
QY 121 YGAPILVDFGTATTCFAISEKGEYLGTTAPGKISSEALFQSASKLPRLVELAK 180
DB 121 YGAPILVDFGTATTCFAISEKGEYLGTTAPGKISSEALFQSASKLPRLVELAK 180
QY 181 CKSTYSAMQSGIYGVGLVDKIISIMKKELNCD 214
DB 181 GRNTYSAMQSGILFGIVGQEGIVARMKWAQKD 214
RESULT 6
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36391
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T36391
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-265 <OLI>
A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCODEB:SCE94.31c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SCE94.31c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 49.3%; Score 643.5; DB 2; Length 265;
Best Local Similarity 49.8%; Pred. No. 7.7e-45;
Matches 130; Conservative 53; Mismatches 71; Indels 7; Gaps 4;
QY 1 MLLVFDVGNNTNVLGIYKDKLVNRYIKTDRKTSDEYGLISNLFYDYNV---NISD- 56
DB 1 MLLTIDVGNTHTVLGFDEGIVEHWRISTDSRRATLAVLQGLGMHPLLDGDLG 60
QY 57 IDDVLISSVVPNMHSLNFCIKYCKKQP-LIVGPGIKTGLNIKYNPKQVADRIYNAV 115
DB 61 IDGIAICATVPSVLHRELVRYTGYDPAVLVEPGVGTVPILTDHPKEVGADRIYNAV 120
QY 116 AGIERYGAPSLVDFGTATTCFAISEKGEYLGTTAPGKISSEALFQSASKLPRLVELAK 175
DB 121 RAVELYGPAIVDGTATTCFAISEKGEYLGTTAPGKISSEALFQSASKLPRLVELAK 180
QY 176 PGMTICKSTVSAMQSGIYGVGLVDKIISIMKKEL--NCDVVKVIATGGLAKLASETK 233
DB 181 PRSVIGKNTVEAMQSGIYGVGAGQVGVNRMARELADDPDVTVTIATGGLAPMLGESS 240
QY 234 SIDYVDGFLTLGLELRIYEKN 254
DB 241 VIDEHEPWLTLMGLRLVVERN 261
RESULT 7
H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H86937
R:Coile, S.T.; Elglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holm
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.,
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H86937
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <STO>
A:Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML0232
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 40.5%; Score 529; DB 2; Length 274;
Best Local Similarity 39.7%; Pred. No. 1.6e-35;
Matches 106; Conservative 63; Mismatches 84; Indels 14; Gaps 4;
QY 1 MLLVFDVGNNTNVLGIYKDKLVNRYIKTDRKTSDEYGLISNLFYDYNVNSID 56
DB 1 MLLAIDVRNTHTVVGLSGSKHAKVYQWRIQTESEVTADELALIDGLIGDSERLA- 59
QY 57 IDDVLISSVVPNMHSLNFCIKYCKKQP-LIVGPGIKTGLNIKYNPKQVADRIYNAV 115
DB 60 --GAALSTVPSVLHRELVRYTGYDPAVLVEPGVGTVPILTDHPKEVGADRIYNAV 117
QY 116 AGIERYGAPSLVDFGTATTCFAISEKGEYLGTTAPGKISSEALFQSASKLPRLVELAK 175
DB 118 AAFHFGQAIVVDFGSSICVDVSAKGEFLGATAPGVQVSSDAAARSALRRVELAR 177
QY 176 PGMTICKSTVSAMQSGIYGVGLVDKIISIMKKELN-----CDDVKVIATGGLAKLTA 229
DB 178 PRSVIGKNTVEAMQSGIYGVGAGQVGVNRMARELADDPDVTVTIATGGLAPMLGESS 240
QY 230 SETKSIDYVDGFLTLGLELRIYEKNQE 256
DB 238 PELHTVDHYDRHLTLHGLRLVERNE 264

RESULT 8

B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: B87489
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolof
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-261 <STO>
A:Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1935
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 40.1%; Score 523.5; DB 2; Length 261;
Best Local Similarity 40.6%; Pred. No. 4.2e-35;
Matches 104; Conservative 53; Mismatches 96; Indels 3; Gaps 2;

QY 1 MLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISD 60
Db 1 MLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISD 60
QY 2 MLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISD 60
Db 2 MLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISD 60
QY 61 IISVVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 120
Db 61 IISVVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 120
QY 62 IISVVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 120
Db 62 IISVVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 120
QY 121 YGAPSLVDGTTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVELAKP--GM 178
Db 121 YGAPSLVDGTTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVELAKP--GM 178
QY 122 YGAPSLVDGTTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVELAKP--GM 178
Db 122 YGAPSLVDGTTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVELAKP--GM 178
QY 179 TICKSTVSAMQSGIIYGVGLVDKIIISIMKELNCDVVIATGGAKLIASFTKSIDYV 238
Db 179 TICKSTVSAMQSGIIYGVGLVDKIIISIMKELNCDVVIATGGAKLIASFTKSIDYV 238
QY 182 IVGTDVTSAMQSGVFGYISLGLVARIKAERG-BPMTVIATGVSALFEGATSDIDHF 240
Db 182 IVGTDVTSAMQSGVFGYISLGLVARIKAERG-BPMTVIATGVSALFEGATSDIDHF 240
QY 239 DGLTLEGLRIIYEKN 254
Db 239 DGLTLEGLRIIYEKN 254
QY 241 DGLTLEGLRIIYEKN 254
Db 241 DGLTLEGLRIIYEKN 254

RESULT 9

A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70955
R:Coale, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70955
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-272 <COL>
A:Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113976
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3600c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 39.3%; Score 513; DB 2; Length 272;
Best Local Similarity 38.9%; Pred. No. 3.2e-34;
Matches 103; Conservative 65; Mismatches 85; Indels 12; Gaps 4;

QY 1 MLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISD 56
Db 1 MLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISD 56
QY 57 IDVVISSVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 115
Db 57 IDVVISSVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 115
QY 60 --GTAALSTVSPVSLVHEVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVGCADRIYNAV 117
Db 60 --GTAALSTVSPVSLVHEVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVGCADRIYNAV 117
QY 116 AGTEKYGAPSLVDGTTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVELAK 175
Db 116 AGTEKYGAPSLVDGTTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVELAK 175
QY 118 AYDRFRKAIYVDFGSSICVDVVSAGFGLGAIAPGVQVSSDAARSAALRRVELAR 177
Db 118 AYDRFRKAIYVDFGSSICVDVVSAGFGLGAIAPGVQVSSDAARSAALRRVELAR 177
QY 176 PGMTICKSTVSAMQSGIIYGVGLVDKIIISIMKELN---CDVVKVIATGGAKLIASE 231
Db 176 PGMTICKSTVSAMQSGIIYGVGLVDKIIISIMKELN---CDVVKVIATGGAKLIASE 231
QY 178 PRSVGKNTVECMQAGAVGAGFAGLVGLVGRIRIEDVSGFSDVDHVAIVATGHTAPLLPE 237
Db 178 PRSVGKNTVECMQAGAVGAGFAGLVGLVGRIRIEDVSGFSDVDHVAIVATGHTAPLLPE 237
QY 232 TKSIDYVDGFLTLEGLRIIYEKNQE 256
Db 232 TKSIDYVDGFLTLEGLRIIYEKNQE 256
QY 238 LHTVDHYDQHLTLQGLRLVFERNLE 262
Db 238 LHTVDHYDQHLTLQGLRLVFERNLE 262

RESULT 10

E75516
conserved hypothetical protein - Deinococcus radiodurans (strain RI)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: E75516
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75516
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <WHI>
A:Cross-references: GB:AE001905; GB:AE005513; NID:g6458144; PIDN:AAF10040.1; PID:g645
A:Experimental source: strain RI
C:Genetics:
A:Gene: DR0461
A:Map position: 1
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.9%; Score 377; DB 2; Length 262;
Best Local Similarity 35.7%; Pred. No. 3.3e-23;
Matches 94; Conservative 51; Mismatches 102; Indels 16; Gaps 9;

QY 2 LLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISDIDDV 60
Db 2 LLVFDVGNNTNMVLYGKGD-KLVNWRRIKTDREKTSDEYGLISNLFYDNDVNISDIDDV 60
QY 61 IISVVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 116
Db 61 IISVVVPVNMVHLENFCIKYCKKQPLIVGPGIKTGLNKNYDNPQVGCADRIYNAV 116
QY 64 VLSSVAPPVGV---ENYALA-LRRHFMDAFVSAENLPDVTVELDPGVSAGRLCN-LF 118
Db 64 VLSSVAPPVGV---ENYALA-LRRHFMDAFVSAENLPDVTVELDPGVSAGRLCN-LF 118
QY 117 GLEKY--GAP--SILVDFGTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVEL 173
Db 117 GLEKY--GAP--SILVDFGTATTCFAISEKEGYLGTTAPGKISSEALFQSKLPRVEL 173
QY 119 GAELKGLGLDYAVVDFGTSTNPDVVGRRFLGGLIATGAQVSADALEFARAKLPRIYL 178
Db 119 GAELKGLGLDYAVVDFGTSTNPDVVGRRFLGGLIATGAQVSADALEFARAKLPRIYL 178
QY 174 AKPGMTICKSTVSAMQSGIIYGVGLVDKIIISIMKELNCDVVIATGGAKLIASETK 233
Db 174 AKPGMTICKSTVSAMQSGIIYGVGLVDKIIISIMKELNCDVVIATGGAKLIASETK 233
QY 179 QAPETAIGKNTVHALQSLGVFGVYAEVMDGLLRIRAEPLGAEV-AVATGGSRTVQIGCQ 237
Db 179 QAPETAIGKNTVHALQSLGVFGVYAEVMDGLLRIRAEPLGAEV-AVATGGSRTVQIGCQ 237
QY 234 SIDYVDGFLTLEGLRIIYEKNQE 256
Db 234 SIDYVDGFLTLEGLRIIYEKNQE 256
QY 238 EIDYDETTLTLEGLVELNASRSE 260
Db 238 EIDYDETTLTLEGLVELNASRSE 260

RESULT 11

D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.14968 Seconds
(without alignments)
2061.866 Million cell updates/sec

Title: US-09-813-453a-55

Perfect score: 1305

Sequence: 1 MLLVFDVGNNTNVLGIYKGD.....YVDGFLTLEGLRIIYKQNE 256

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	672	51.5	233	1	YACB_BACSU	P37564 bacillus su
2	127.5	9.8	267	1	BAF_BORPE	Q45338 bordetella
3	105	8.0	1159	1	DP3A_PASMU	Q9cpk3 pasteurella
4	102.5	7.9	1159	1	DP3A_HAEIN	P43743 haemophilus
5	95.5	7.3	513	1	ASPR_CUCPE	O04057 cucurbita p
6	95	7.3	1161	1	DP3A_BUCAI	P57332 buchnera ap
7	94	7.2	1067	1	CARB_CLOPE	Q8xnb3 clostridium
8	92.5	7.1	381	1	ADHX_MAIZE	P93629 zea mays (m
9	92.5	7.1	1036	1	SYI_CHLMU	Q9p120 chlamydia m
10	91.5	7.0	381	1	ADHX_ORYSA	P93436 oryza sativ
11	89.5	6.9	735	1	DHB4_RAT	P97852 rattus norv
12	89.5	6.9	1173	1	DP3A_PSEFL	Q9xhb6 pseudomonas
13	89	6.8	393	1	LOLC_HAEIN	P42552 haemophilus
14	88	6.7	553	1	VGLF_NDVU	P12570 newcastle d
15	87.5	6.7	247	1	DP33_MYCPN	P75164 mycoplasma
16	87.5	6.7	1159	1	DP3A_VIBCH	P52022 vibrio chol
17	86.5	6.6	303	1	V212_FOWPV	Q91523 fowlpox vir
18	86.5	6.6	1173	1	DP3A_PSEAE	Q9hxz1 pseudomonas
19	86	6.6	304	1	HEM3_AQUAE	O66621 aquifex aeo
20	86	6.6	443	1	YK24_YEAST	P33137 penicillium
21	85.5	6.6	417	1	PKK_PENCI	P33161 penicillium
22	85.5	6.6	623	1	PRR2_CANAL	P46030 candida alb
23	85.5	6.6	1160	1	DP3A_SALTY	P14567 salmonella
24	84.5	6.5	378	1	ADHX_PEA	P80572 pisum sativ
25	84.5	6.5	501	1	YH92_CAEEL	Q23256 caenorhabdi
26	84.5	6.5	735	1	DHB4_MOUSE	P51660 mus musculu
27	84.5	6.5	1160	1	DP3A_ECOLI	P10443 escherichia
28	84	6.4	553	1	VGLF_NDVU	P35936 newcastle d
29	84	6.4	587	1	RUBA_PEA	P08926 pisum sativ
30	84	6.4	725	1	PALY_TRISU	P45734 trifolium s
31	83.5	6.4	335	1	GCP_AQUAE	O66986 aquifex aeo
32	83.5	6.4	591	1	SYR_BORBU	O51540 borrelia bu
33	83.5	6.4	1115	1	DP3A_BACSU	O34623 bacillus su

34	83	6.4	190	1	PYRE_THETN	P58858 thermoanaer
35	83	6.4	314	1	MIAA_BACSU	O31795 bacillus su
36	83	6.4	368	1	ISPG_BUCAI	P57374 buchnera ap
37	83	6.4	467	1	LEO2_VIBCH	Q9xp81 vibrio chol
38	83	6.4	553	1	VGLF_NDVH4	P33613 newcastle d
39	83	6.4	558	1	NADE_SYNY3	P74292 synecocyst
40	83	6.4	991	1	SCA4_RICAU	Q9aj64 rickettsia
41	82.5	6.3	425	1	GSA_CLOPE	Q9znc8 clostridium
42	82.5	6.3	452	1	MRJ2_APIME	O77061 apis mellif
43	82.5	6.3	969	1	DF3A_UREPA	Q9pq74 ureaplasma
44	82	6.3	206	1	UREG_UREPA	Q56561 ureaplasma
45	82	6.3	286	1	AROE_LACLA	Q9ces7 lactococcus

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D26185; BAA05305.1; -
 CC EMBL; 299104; CAB11846.1; -
 CC Subtilist; BG10133; YACB.
 CC InterPro; IPR004619; Baf.
 CC Pfam; PF03309; Bvg_acc_factor; 1.
 CC TIGRFAMS; TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;
 Query Match 51.5%; Score 672; DB 1; Length 233;
 Best Local Similarity 60.3%; Pred. No. 5.2e-45;
 Matches 129; Conservative 32; Mismatches 53; Indels 0; Gaps 0;
 Y 1 MLVFDVGNNTVGLYKDKLVNRYRIKTDREKTSDEYGLILSNLFDVNDVNSIDDDV 60
 Db 1 MLVFDVGNNTVGLYHDKLEYHWRITSRHKTEDEFGMLRSLDFHSGLMFEQIDGI 60
 QY 61 IISVVVNVNHSLENFCIKYCKKQPLIVGPGIKTGLNRYDNPQVGDRIVNAVAGIEK 120
 Db 61 IISVVVVPINFALERCTYFHEPQIVGPGMTGLNRYDNPQVGDRIVNAVAAIHL 120
 QY 121 YGAPSLVDFGTATTCATSEKGEYLGTTAPGKISSEALFOSAKLPRVELAKPGMTI 180
 Db 121 YGNPLVDFGTATTCYIDENQYMGGAIPGTTISTEALYSRAAKLPRIETRPDNI 180
 QY 181 CKSTVSAMQSGLIYGVGLVDKIISTMKKELNCD 214
 Db 181 GKNTVSAMQSGLIYGVGVGIVKRWQAKOD 214

RESULT 2

BAF_BORPE STANDARD; PRT; 267 AA.
 AC Q45373; Q45373;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Bvg accessory factor.
 GN BAF.
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 RX Bordetella.
 NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP504;
 RX MEDLINE=95325323; PubMed=7601846;
 RA Deshazer D., Wood G.E., Friedman R.L.;
 RT Identification of a Bordetella pertussis regulatory factor required
 RT for transcription of the pertussis toxin operon in Escherichia
 coli.;
 RL J. Bacteriol. 177:3801-3807(1995).
 RN [2]
 RP SEQUENCE OF 1-38 FROM N.A.
 RC STRAIN=BP504;
 RA Wood G.E., Friedman R.L.;
 RT Identification of a bira homolog in Bordetella pertussis.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 239-267 FROM N.A.
 RC STRAIN=BP536;
 RX MEDLINE=96419162; PubMed=8821935;
 RA Allen A.G., Maskell D.J.;
 RT The identification, cloning and mutagenesis of a genetic locus

RT required for lipopolysaccharide biosynthesis in Bordetella
 RT pertussis.;
 RL Mol. Microbiol. 19:37-52(1996).
 CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
 CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
 CC RNA POLYMERASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U12020; AAA75361.1; -
 CC EMBL; AF016461; AAC68834.1; -
 CC EMBL; X90711; CAA62242.1; -
 CC InterPro; IPR004619; Baf.
 CC Pfam; PF03309; Bvg_acc_factor; 1.
 KW Transcription regulation; Activator.
 SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;
 Query Match 9.8%; Score 127.5; DB 1; Length 267;
 Best Local Similarity 31.7%; Pred. No. 0.006;
 Matches 40; Conservative 17; Mismatches 58; Indels 11; Gaps 4;
 QY 78 IYCKKQPLIVGPGIKTGLNRYDNPQVGDRIVNAVAGIEKYPSPILVDFGTAT 134
 Db 80 IRLWRAQPLAM--GLRNG----YRNPDLGLGACWACWAGVLRQPSVHPPLVASEGTAT 133
 QY 135 TFCATSEKGEYLGTTAPGKISSEALFOSAKLPRVELAKPGMTICKSTVSAMQSGLIY 194
 Db 134 TLDITGPDNVFPGILPQPMRGALATGTAHLPLADGLVADYPI--DTHQAIASGIAA 191
 QY 195 GYVGLV 200
 Db 192 AQAGAI 197
 RESULT 3
 DP3A_PASMU STANDARD; PRT; 1159 AA.
 AC Q9CPR3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA polymerase III alpha subunit (EC 2.7.7.7).
 GN DNAE OR PM0034.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 RX Pasteurella.
 NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT Complete genomic sequence of Pasteurella multocida PM70.;
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate
 CC + [DNA](N).
 CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
 CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
 CC DIMERIZATION TO FORM THE POLIIT' COMPLEX. POLIIT' ASSOCIATES WITH
 CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
 CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE

[1] SEQUENCE FROM N.A.
STRAIN-Rd / KW20 / ATCC 51907;
MEDLINE=95350630; PubMed=7542800;
Fleischmann R.D., Adams M.D., Clayton R.A., Kirkness E.F.,
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
Fine L.D., Fritchman J.L., Fuhrmann J.L., Geochagen N.S.M.,


```
EMBL; AP003194; BAB82278.1; -.
InterPro; IPR005483; CPase_L.
InterPro; IPR005479; CPase_L_D2.
InterPro; IPR005480; CPase_L_D3.
InterPro; IPR005481; CPase_L_N.
InterPro; IPR004362; MGS_like.
Pfam; PF02789; CPsase_L_chain; 2.
Pfam; PF02786; CPsase_L_D2; 2.
Pfam; PF02787; CPsase_L_D3; 1.
Pfam; PF02142; MGS; 1.
PRINTS; PR00098; CPsase.
PROSITE; PS00866; CPsase_1; 2.
PROSITE; PS00867; CPsase_2; 2.
Arginine biosynthesis; Pyrimidine biosynthesis; Liase; Repeat;
```

KW	ATP-binding; Manganese; Complete protosome.						
FT	DOMAIN	1	401	CARBOXYPHOSPHATE SYNTHETIC DOMAIN.			
FT	DOMAIN	402	598	OLIGOMERIZATION DOMAIN.			
FT	DOMAIN	550	932	CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.			
FT	DOMAIN	933	1067	ALLOSTERIC DOMAIN.			
FT	REPEAT	1	549	:			
FT	REPEAT	550	1067	:			
FT	NP_BIND	153	210	ATP (POTENTIAL).			
FT	NP_BIND	302	352	ATP (POTENTIAL).			
FT	METAL	284	284	MANGANESE 1 (BY SIMILARITY).			
FT	METAL	298	298	MANGANESE 1 AND 2 (BY SIMILARITY).			
FT	METAL	300	300	MANGANESE 2 (BY SIMILARITY).			
FT	METAL	823	823	MANGANESE 3 (BY SIMILARITY).			
FT	METAL	835	835	MANGANESE 3 (BY SIMILARITY).			
SQ	SEQUENCE	1067 AA;	118897 MW;	A4C2E518224611A6 CRC64;			
 Query Match 7.2%; Score 94; DB 1; Length 1067; Best Local Similarity 23.9%; Pred.No. 11; Matches 64; Conservative 39; Mismatches 79; Indels 86; Gaps 15							
QY	33	EKTSDEYGI--LISNLFDDYNV-	-----ISDIDDIIVSSV----	66			
Dd	721	EITHDEELTYLTNAFEDSKNPILD	KYLMGRETEVDASDGEDVLPGIMEHLE	RAG 780			
QY	67	-----PNVMHSLENFCIKCKKKOPL	IIVGGCIKTGLNIKY-----DNP	103			
Dd	781	VHSGDSITMYPQNISKDKIEDVLD	YTTRKLALSII--GIRKMNIQIIEFG	KLYVIEVP 838			
QY	104	KQVGADRIYNAGIEKIYCAPSLVD	FGTATTFCAISEKEYLIG--GTIAPI	KISSSEA 160			
Dd	839	R---ASRTVPYISKVS--GVPT--	IVDIATR---IMLGKLLDLGYGTGVKE	PCLYSVKV 888			
QY	161	LFSQASKLPVELLA-KPGMT-----	ICKSTSAMSQSGLIYGVLVDK---II	SIIMKR 209			
Dd	889	PVFESTOKLVNEVSVLSGPMSRST	GCVLGVGRNFVEALKGFVGASMVTGD	KTILAIKK 948			
QY	210	-----ELNCDDVKV-----IAT	GLAK 226				
Dd	949	HDKEFMELAKDLKLGYNFIATTGT	AK 976				
 RESULT 8 ADHX_MAIZE STANDARD; PRT; 381 AA. AC P93629; DT DT 30-MAY-2000 (Rel. 39, Created) DT DT 30-MAY-2000 (Rel. 39, Last sequence update) DT DE 30-MAY-2000 (Rel. 39, Last annotation update) DE DE Alcohol dehydrogenase class III (EC 1.1.1.1) (Glutathione-dependent FDH. formaldehyde dehydrogenase) (EC 1.2.1.1) (FALDH) (GSH-FDH). OS Zea mays (Maize). OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACC clade; OC Panicoideae; Andropogoneae; Zea. OX NCBI_TaxId=4577; RN [1] RP SEQUENCE FROM N.A. RC STRAIN=cv, Black Mexican Sweet; RX MEDLINE=J7435970; PubMed=9290637; RA Fliegmann J., Sander mann H. Jr.; RT "Maize glutathione-dependent formalde hyde dehydrogenase cDNA: a novel RT plant gene of detoxification."; RL Plant Mol. Biol. 34:843-854(1997). CC - - CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone + CC NADH. CC - - CATALYTIC ACTIVITY: Formaldehyde + glutathione + NAD(+) = S- CC formylglutathione + NADH. CC - - COFACTOR: ZINC (BY SIMILARITY). CC - - SUBCELLULAR LOCATION: Cytoplasmic (Potential); CC - - TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN THE LEAVES. CC - - SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE CC FAMILY. CLASS:-III SUBFAMILY.							


```

CC + [DNA](N).
CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLIIV' COMPLEX. POLIIV' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, PSI,
CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF062919; AD44402.1; -.
CC InterPro: IPR004013; PHP_C.
CC InterPro: IPR003141; PHP_N.
CC InterPro: IPR004805; PolC_alpha.
CC InterPro: IPR004365; tRNA_anti.
CC Pfam: PF01336; tRNA_anti; 1.
CC Pfam: PF02231; PHP_N; 1.
CC Pfam: PF02811; PHP_C; 1.
CC SMART: SM00481; POLIIVAC; 1.
CC TIGRFAMs: TIGR00594; polc; 1.
CC Transferase: DNA-directed DNA polymerase: DNA replication.
CC KW SQUENCE 1173 AA; 131216 MW; E0BF3158F3876E55 CRC64;
CC -----
Query Match 6.9%; Score 89.5; DB 1; Length 1173;
Best Local Similarity 21.1%; Pred. No. 28;
Matches 48; Conservative 39; Mismatches 80; Indels 61; Gaps 9;
CC QY 59 DVIISVVPVNHSLNFCIKYCKKQPLIVGPIKTLN-----IKYD--- 101
CC Db 332 DIIIQMGFGYFLIVMDF-IQAKNNGVPGPGRGSGAGSLVAYVQKITDLPLEYDLLF 390
CC QY 102 -----NPKQV-----GADRVNAVAGIEKYG--APSTLVDFGTATTCATSEKG 143
CC Db 391 ERLNPERVSMPPDFVDFCMGDRVIDYA--EKYGRNAVSGIITFGSMAKAVRDVA 448
CC QY 144 EYLGGTIAPGKITSBALFQSKLPRVELAKPGMTICKSTVSAMQSGIYGVGLVDKI 203
CC Db 449 RAQCKS-----FGLADRLSKMVPFEVGMTLEK---AYEQEILRDFIKVDEEA 493
CC QY 204 ISIMKELNCDYKVVIATGGLAKLIASETKSIDYV-----DGFLT 243
CC 494 AEIEMARKLEGVVRNVGKHAGGVIAPTKLTDFSPIYCDAGDGLVT 541
CC -----
RESULT 13
LOLC_HAEN
ID LOLC_HAEN STANDARD; PRT; 393 AA.
AC P44252;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein releasing system transmembrane protein loic.
GN LOIC OR HI1555.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rg / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

```

```

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: PART OF AN ATP-DEPENDENT TRANSPORT SYSTEM LOICDE
CC RESPONSIBLE FOR THE RELEASE OF LIPOPROTEINS TARGETED TO THE OUTER
CC MEMBRANE FROM THE INNER MEMBRANE. SUCH A RELEASE IS DEPENDENT OF
CC THE SORTING-SIGNAL (ABSENCE OF AN ASP AT POSITION 2 OF THE MATURE
CC LIPOPROTEIN) AND OF LOLA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (BY similarity).
CC -1- SIMILARITY: BELONGS TO THE LOIC/E FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32830; AAC23204.1; -.
CC TIGR: HI1555; -.
CC InterPro: IPR003838; DUF214.
CC Pfam: PF02687; DUF214; 1.
CC Transport; Transmembrane; Inner membrane; Complete proteome.
CC TRANSMEM 25 45 POTENTIAL.
CC TRANSMEM 262 282 POTENTIAL.
CC TRANSMEM 314 334 POTENTIAL.
CC TRANSMEM 356 376 POTENTIAL.
CC SEQUENCE 393 AA; 43579 MW; 7ABEC4E335D57F34 CRC64;
CC -----
Query Match 6.8%; Score 89; DB 1; Length 393;
Best Local Similarity 22.6%; Pred. No. 8.7;
Matches 51; Conservative 32; Mismatches 67; Indels 76; Gaps 12;
CC QY 11 NNVILGIYKDKLVNRYRIKTRDEKTSDEYGLISNLFYDNVNISIDDDVISSVVPNV 70
CC Db 2 NFPISLYIA---LRYW-----RAKSADREFGLVTNL-----ASLGIVLGYMALIIVLSV 48
CC QY 71 HSLNFCIKYCKKQPLIVG-----GIKTLNLIKVDNP 103
CC Db 49 NLEG-----YQKQVLSIPHAIVSEEQIPSTEKTLENLPHFVOKAVPINT-TNVIQTA 103
CC QY 104 KQVGADRIYNVAGIEKYGAPSLVDFGTATTCATSEKGEY-----LGGTIAP 152
CC Db 104 KGVSAQII---GIQSFSDPLVESF-DQTKFNEILPRGEFLVIGDQLAOKLGVNIGD 158
CC QY 153 GIK--ISSEALFQSKLPRVELAKPGMTICKSTVSAMQSGIYGY 196
CC Db 159 KIRLMITENSQYTPFGRVPMQRL-----FTV-----SDIYGY 191
CC -----
RESULT 14
VGLF_NDVU
ID VGLF_NDVU STANDARD; PRT; 553 AA.
AC P12570;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2;
DE Fusion glycoprotein F1].
GN F.
OS Newcastle disease virus (strain Ulster/67) (NDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX NCBI_TaxID=11190;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=88171450; PubMed=3351479;
RA Millar N.S., Chambers P., Emerson P.T.;
RT "Nucleotide sequence of the fusion and haemagglutinin-neuraminidase
RT glycoprotein genes of Newcastle disease virus, strain Ulster";
RT molecular basis for variations in pathogenicity between strains.";
RL J. Gen. Virol. 69:613-620(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204898; PubMed=2705298;
RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
RA Nagai Y.;
RT "Newcastle disease virus evolution. II. Lack of gene recombination in
RT generating virulent and avirulent strains.";
RL Virology 169:273-282(1989).
CC -!- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
CC MEMBRANES.
CC -!- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D00243; BAA00173.1; -
DR EMBL; M24694; AAA46645.1; -
DR PIR; A29823; VGN201.
DR InterPro; C46329.
DR Pfam; PF00523; fusion_gly; 1.
KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
FT SIGNAL 1 25
FT CHAIN 26 553 FUSION GLYCOPROTEIN F0.
FT CHAIN 26 116 F2 PROTEIN.
FT CHAIN 117 553 F1 PROTEIN.
FT TRANSMEM 117 136
FT DOMAIN 137 500 POTENTIAL.
FT TRANSMEM 501 527 EXTRACELLULAR.
FT POTENTIAL.
FT CYTOPLASMIC.
FT DOMAIN 528 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 523 PALMITATE (POTENTIAL).
SQ SEQUENCE 553 AA; 58682 MW; FC99C763D8E2F7FA CRC64;

Query Match 6.7%; Score 88; DB 1; Length 553;
Best Local Similarity 21.4%; Pred. No. 15;
Matches 72; Conservative 52; Mismatches 117; Indels 96; Gaps 17;

QY 8 GNTNVLGIYKDKLVNRYWRIKTDRE-----NTSDEYGLISNLF--DYDNVNISDIDV 61
DB 171 GLSLQAVAGKMQQQVNDQFNKTAQELDCIKITQGVGVEL-NLYLTELTTVFGQITSPA 229
QY 62 ISSVVPVNMHSLNFCIKYK-----KKOPLVGPQIKTGLNKNKPNQVQADRI-- 111
DB 230 LTQTLQIALYNLAGNDMDYLLTLKGVGNQNSLLIGSLGILGNPILYDSQTLGIQVTL 289
QY 112 -----VNAGIEKYGAPSLVDFGTA-----TTFCAISEKGEYLG 148
DB 290 PSVGNNLNRRATYLETTSVSTTKGFASALVPKVVTVQGVSEELDTSYCIETDLDTCT 349
QY 149 TI-----APGIKISSEALFQASAKLPVRE--LAKFGWTF-----CKSTV--SAMQSGII 193
DB 350 IVTFPMSPGI-YSLSGNTSACMYSKTEGALTTPYMTLKGSVIANCKMTTCRSADPPGII 408

QY 194 ---YG-YVGLVDK-----IISI-----MKKELNDDVKVIATG----- 222
DB 409 SONYGEANSLDRQSCNVLSLDCIYILRSGEFDAYQKNISQDSOVITGNLDISTELG 468
QY 223 -----GLAKLIASETKSIDYVDGFLTFLEGLRIY 251
DB 469 NVNNSISNALDKLEESNSK-LDKVNVKLTSTSTALITY 504
RESULT 15
YQ33_MYCPN
ID YG33_MYCPN STANDARD; PRT; 247 AA.
AC P75164;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN633 (C12_Orf247).
GN MPN633 OR MP209.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: STRONG, TO THE N-TERMINAL OF M.PNEUMONIAE MPN635.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF000021; AAB95857.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 247 AA; 28095 MW; B9C82D8F1B38CEFE CRC64;

Query Match 6.7%; Score 87.5; DB 1; Length 247;
Best Local Similarity 24.3%; Pred. No. 6.6;
Matches 54; Conservative 38; Mismatches 85; Indels 45; Gaps 11;

QY 9 NTNMYL-----GIYKDKLVNRYWRIKTDREKTSDEYGI-----LISNLFYDNVNISDIDD 59
DB 50 NSSLVKIDYGRGKSNHFTQNESREKQVSEKTIKFGIGLKDAIYVLPFH-NVKYSFTSS 108
QY 60 VIISVVPVNMHSLNFCIKYKCKOPLVGPQIKTGLNKNKPNQV--GADRIYNAVAG 117
DB 109 EGTFTPVERMKEGMD-----GTTK-IQITVDETRKIDKGTDLISKISR 152
QY 118 LEKYCAPSLVDVDFGATTFCAISEKGEYL-----GGTI-APGIKISSEA--LFQASAKLP 170
DB 153 SDYKAIATFLRLTGYOKLASSKKGDIYRSENGSEIFLNGMKITGDFLFSYDIKEPN 212
QY 171 VELAKPGMTCKTSTVSAMQSGIYGVGLVDKRIISIMKKELN 212
DB 213 KKLR-SLNREKTLR-----DSVRDNIISILKSSIN 244

Search completed: June 24, 2003, 22:11:53
Job time : 7.14968 secs

THIS PAGE BLANK (USPTO)

[illegible]

```

Db      240 FLTLKGLLELYRRNK 254
RESULT 7
Q92F54  PRELIMINARY; PRT; 259 AA.
AC      Q92F54
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Hypothetical protein lin0253.
GN      Listeria innocua.
OS      Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC      Listeriaceae; Listeria.
OX      NCBI_TaxID=1642;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CLIP 11262 / SROVAR 6A;
RX      PubMed=11679669;
RA      Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA      Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
A      Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA      Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA      Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA      Gautier L., Gobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA      Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA      Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA      Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA      Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA      Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT      "Comparative genomics of Listeria species.";
RL      Science 294:849-852(2001).
DR      EMBL; AL596164; CAC95486.1; -
DR      ListList; LIN00253; -
DR      InterPro; IPR004619; Baf.
DR      Pfam; PF03309; Bvg_acc_factor; 1.
DR      TIGRFAMS; TIGR00671; baf; 1.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 59.6%; Score 777.5; DB 16; Length 259;
Best Local Similarity 56.5%; Pred. No. 2.5e-55;
Matches 144; Conservative 51; Mismatches 59; Indels 1; Gaps 1;

QY      1 MLIVFDVGNNTMVLGIYKGDKNVYRIKTDREKTSDEYGLISNLFYDNNISIDVV 60
Db      1 MLIVDVGNTCTGVYKQKLLRHWRMTDRHTSDELGMTVLNFFSYANLTPSDIQI 60
61 IISVVVPNMVHSLNFCIKYCKQPLIVGPIKTNLNIYDNPQVAGDRIVNAVAGIEK 120
61 IISVVVPIMHAMETWCVRVFNRLPLIVGPIKTNLNLKVDNPNREIGSDRIVNAVASEE 120
QY      121 YGAPSLVDFGTATTCAISEKGEYLGGTIAPGKISSEALFQSKLPRVELAKPGWI 180
Db      121 YGPIVDFGTATTCYIDEAGVYQGAIPAGIMISTEALYNRAAKLPVDIAESQII 180
QY      181 CKSTVSAMSGIYGVGVLDKIISIMKKELNDDVKVIATGATGLAKLIASETKSIDYVG 240
Db      181 GKSTVSMQAGIFYGIGQCEGIIAEMKKQSNTPSPV-VVATGGLARMITEKSSAVDILDP 239
QY      241 FLTLEGLRIIYEKN 255
Db      240 FLTLKGLLELYRRNK 254

RESULT 8
Q9X8N6  PRELIMINARY; PRT; 265 AA.
AC      Q9X8N6
ID      Q9X8N6
DT      01-NOV-1999 (TrEMBLrel. 12, Created)
DT      01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE      Hypothetical protein SCO3380.
GN      SCO3380 OR SCE94.31C.
OS      Streptomyces coelicolor.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX      NCBI_raxid=1902;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2);
RA      Oliver K., Harris D.;
RL      Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2);
RA      Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL      Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2);
RX      MEDLINE=97000351; PubMed=8843436;
RA      Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RA      Kinashi H., Hopwood D.A.;
RT      "A set of ordered cosmids and a detailed genetic and physical map for
RL      the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A3(2) / M145;
RA      Thomson S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA      Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA      Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA      Huang C.-H., Kieser T., Larke L., Larke L., Murphy K., O'Neill S.,
RA      Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA      Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA      Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA      Hopwood D.A.;
RT      "Complete genome sequence of the model actinomycete Streptomyces
RL      coelicolor A3(2).";
RL      Nature 417:141-147(2002).
DR      EMBL; AL049628; CAB40880.1; -
DR      InterPro; IPR004619; Baf.
DR      Pfam; PF03309; Bvg_acc_factor; 1.
DR      TIGRFAMS; TIGR00671; baf; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 49.3%; Score 643.5; DB 16; Length 265;
Best Local Similarity 49.8%; Pred. No. 2e-44;
Matches 130; Conservative 53; Mismatches 71; Indels 7; Gaps 4;

QY      1 MLIVFDVGNNTMVLGIYKGDKNVYRIKTDREKTSDEYGLISNLFYDNNV--NISD- 56
Db      1 MLITDVGNTHTVLGDFGDIHWRISTDSRTADELAVLQGLMGHPLGLGDELGDG 60
QY      57 IDDVIISSVVPNMVHSLNFCIKYCKQKP-LIIVGPIKTNLNIYDNPQVAGDRIVNAV 115
Db      61 IDGIAICATVPVSLHRELVRTRYRGDPAVLVEPGVKGTGVPILTDHPKEVGADRIINAV 120
QY      116 AGIEKYGAPSLVDFGTATTTCFAISEKGEYLGGTIAPGKISSEALFQSKLPRVELAK 175
Db      121 AAVELYGGPAIVVDFGTATTTFDAVSARGEYIGGVIAPIGIEISVEALGVGAQURKIEVAR 180
QY      176 PGMTCICKSTVSAMSGIYGVGVLDKIISIMKKEL--NCDVDVKVIATGGLAKLIASETK 233
Db      181 PRSVIGKNTVEAMQSGIVYGFAGQVGVVNRMARLADDDPDVTVIATGGLAPVWLGESS 240
QY      234 SIDYVDGFLTLEGLRIIYEKN 254
Db      241 VIDEHEPWLTLMLGLRLVYERN 261

RESULT 9

```



```
Q9CD56
ID Q9CD56 PRELIMINARY; PRT; 274 AA.
AC Q9CD56;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein ML0232.
GN ML0232.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC Actinomycetales; Corynebacteriineae; Mycobacteriineae; Mycobacterium.
OX NCBI_TaxID=1769;
RN NCBI_TaxID=1769;
RT "Complete genome sequence of Caulobacter crescentus."
SQ SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
DR EMBL: AL583917; CAC29740.1;
DR Leproma; ML0232;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 40.5%; Score 529; DB 16; Length 274;
Best Local Similarity 39.7%; Pred. No. 4.2e-35;
Matches 106; Conservative 63; Mismatches 84; Indels 14; Gaps 4;

QY 1 MLLVFDVGNNTNVLGIYKGD---KLVNYWRKTDREKTSDEYGLISLNFYDYNVNSID 56
DB 1 MLLAIDVNTHTVGLSSKEHAKVVOOVRIRTESEVTADELALIIDGLIGDSSERLA- 59
QY 57 IDDVIISSVNPVNMHSLNFCIKYKQKOP-LIVGPKIKTKGNIKYDNPQKQVADRIVNAV 115
DB 60 --GAAALSTVPSVLHEVRIMLDQYWPSPVPHVLIIEPVGFTGPLLVNDNKEVGADRIVNCL 117
Y 116 AGIEKYGPSILVDPGTATTCATSEKGEYLGTTIAPGKISSEALFQSAKLPRVELAK 175
DB 118 AAFHFGQAAIVDFGSSICVDVVSAGKEFLGGAIPGVQVSSDAAAARSAAALRRVELAR 177
QY 176 PGMTCCKSTVSAMQSGIYGVGLVDKIISIMKKELN-----CDDVKVIATGGLAKLIA 229
DB 178 PRSVGKNTVECMQAGVVGAGLVGLVGRMRQDVEEFGDLGNRVAVATGHTAPLL 237
QY 230 SFTKSIDYVDGFLTLEGLRIIYEKNQE 256
DB 238 PELHTVDVHDLTLHLGLRLYFERNRE 264

RESULT 10
Q9AG21
ID Q9AG21 PRELIMINARY; PRT; 261 AA.
AC Q9AG21;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
```

```
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feidblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005867; AAK23910.1;
DR TIGR: CC1935;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 40.1%; Score 523.5; DB 16; Length 261;
Best Local Similarity 40.6%; Pred. No. 1.1e-34;
Matches 104; Conservative 53; Mismatches 96; Indels 3; Gaps 2;

QY 1 MLLVFDVGNNTNVLGIYKGDKLVNYWRKTDREKTSDEYGLISLNFYDYNVNSIDDD 60
DB 2 MLLAIEOGNTNMTFAHDGASWVAQWRSATSTRTADEYVWLSQLLSMQGLGFRADAV 61
QY 61 IISVVVNPVNMHSLNFCIKYKQKPLIVGPGIKTGLNLIKYNKQVQKQVADRIVNAVAGIEK 120
DB 62 IISVVVQPSIENLRNLRRYFNVEPLVIGENAKLGDIVRIEKPSEAGADRLVNAIGAAVV 121
QY 121 YGAPSLVDVDFGTTTFCATSEKGEYLGTTIAPGKISSEALFQSAKLPRVELAKP--GM 178
DB 122 YGPLVWIDSGTATTFDVAADGAFEGGIAPGILNLSMQALHEAAKLPRIAIQRPAGNR 181
QY 179 TICKSTVSAMQSGIYGVGLVDKIISIMKKELNCDVKVIATGGLAKLIASTESIDV 238
DB 182 IVGTTVSAMQSGVFWGISLIEGLVARIKAERG-EPMTVIATGGVASLFEGETSIDHF 240
QY 239 DGFLFLEGLRIIYEKN 254
DB 241 DSDLTIRGLLEIYRN 256

RESULT 11
O06282
ID O06282 PRELIMINARY; PRT; 272 AA.
AC O06282;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative, Baf family).
GN RV3600C OR MTCY07H7B.22 OR MT3706.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC Actinomycetales; Corynebacteriineae; Mycobacteriineae; Mycobacterium.
OX NCBI_TaxID=1773;
RN NCBI_TaxID=1773;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
```

complete genome sequence.";
Nature 393:537-544(1998).

[2]

SEQUENCE FROM N.A.

STRAIN-CDC 1551 / OSHKOSH;

Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

Bishai W.

"Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains."

Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

EMBL; Z95557; CAB08944.1; -

EMBL; AE007170; AAK48063.1; -

TIGR; MT3706; -

TubercuList; Rv3600c; -

InterPro; IPR004619; Baf.

Pfam; PF03309; Bvg_acc_factor; 1.

TIGRFAMS; TIGR00671; baf; 1.

Hypothetical protein; Complete proteome.

SEQUENCE 272 AA; 29304 MW; 5D70E6EF0F09AC8B CRC64;

Query Match 39.3%; Score 513; DB 16; Length 272;

Best Local Similarity 38.9%; Pred. No. 8.4e-34;

Matches 103; Conservative 65; Mismatches 85; Indels 12; Gaps 4;

1 MLVFDVGNMVLGIYKGD-KLVNRYRIKTDREKTSDEYGLISNLFDDYDNVNSID 56

1 MLVFDVGNMVLGIYKGD-KLVNRYRIKTDREKTSDEYGLISNLFDDYDNVNSID 56

57 IDVIVSSVPMVHSLNFCIKYKQP-LIIVPGIKTGLNLYKYNPKQVQADRVNAV 115

60 -GTAALSTVPSVLHEVRIMLDQWSPVHPVLEPGVGTGIPLLVDNPREVGADRVNCL 117

116 AGTEKCAPSILVDFGTATTCFAISEKGEYLGITAPGKISSEALFQSKLPVELAK 175

118 AAYDRFKAIVVDFGSSICVDVVSAGKFLGAIAPGVQVSDAAARSALRRVELAR 177

176 PGMTICKSTVMSQGIIVGVLDKIIISIMKELN---CDVVKVIATGGLAKLIASE 231

178 PRSVGKNTVECMQAGVAGFAGLVGLVGRIRDSVSDHVDVAIVATGHTAPILLPE 237

232 TKSIDYVDGFLTEGLRIIYKNOE 256

238 LHTVDVHQHLLTGLRLVFNLE 262

RESULT 12

Q9RFE4 PRELIMINARY; PRT; 256 AA.

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

Q9RFE4

```
QY 61 IISVVVPMHSLNFCIKYCKKOPLIYVPGIKT-----GLNIKYDNPQVQVADRIYNAVA 116
DB 64 VLSSVAPPVG---ENVALA-LKRHEMIDAFSAENLPDVTVELDTPGSGADRLCN-LF 118
QY 117 GLEKY--CAP-SILVDFGTATTCFCAISEKGEYLGTTIAPGKISSEALFQSAKSLPRVEL 173
DB 119 GAERYLGLDVAVVVDFGTSNFDVVGRRRFLGILATGAQVSADALFARAAKLPRIITL 178
QY 174 AKPGMTICKSTVSAMOSGIYGVGLVDKIISIMKKELNCCDDVKVIATGGGLAKLIASETK 233
DB 179 QAPETAIGNKTNVHALQSLGVGYAEVMDGLLRIRAEPLPGEAV-AVATGGGSRVTQVGIQ 237
QY 234 SIDYVDGFLTLEGLRIIEKNOE 256
DB 238 EIDYYDETTLRGLVWELNARSSE 260

RESULT 14
ID O83446 PRELIMINARY: PRT: 273 AA.
AC O83446;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein TP0431.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback J.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC65417.1; -.
DR TIGR; TP0431; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 27.1%; Score 354; DB 16; Length 273;
Best Local Similarity 34.1%; Pred. No. 7e-21; Mismatches 99; Indels 14; Gaps 6;
Matches 87; Conservative 55;

QY 2 LIAVDFVGNMVLGIYKDK-----LVNWRKTDRKTSDEYGLISLNFYDNNISD 56
DB 1 MLIDVGNHVVFGI-QGNGRVCVRFLAPARKTDQDEYSLIHALCERAGVGRAS 59
QY 57 IDDVIISSVVPNMHSLNFCIKYCKKOPLIYVPGIKTGLNIKYDNP--KQVGADRIYNA 114
DB 60 LRDAFISVVVPLTKTIADAVAQISGVQPVVFGPWAYEHLVPRIPEVRAEIGTDLVANA 119
QY 115 VAGIEKYGAPSLVDGFTATTCFCAISEKGEYLGTTIAPGKISSEALFQSAKSLPRVELA 174
DB 120 VAAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLA 179
QY 175 KPGMTICKSTVSAMOSGIYGVGLVDKIISIMKKELNCCDDVKVIATGGGLAKLIASETKS 234
DB 180 LPDSVLGKDTTHAVAGVVRGTLFVIRAMIAQCQKELGRCRAAVI-TGGLSLRFSSE--- 235
QY 235 IDY--VDGFLTLEGL 247
DB 238 EIDYYDETTLRGLVWELNARSSE 260
```

```
DB 236 VDFPPIDAQLTLSGL 250

RESULT 15
QY 09WZY5 PRELIMINARY: PRT: 246 AA.
AC 09WZY5;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -.
DR TIGR; TM0883; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 24.6%; Score 321; DB 16; Length 246;
Best Local Similarity 34.4%; Pred. No. 2.9e-18;
Matches 86; Conservative 45; Mismatches 105; Indels 14; Gaps 7;

QY 1 MLLVFDVGNMVLGIYKDKLVNWRKTDRKTSDEYGLISLNFYDNNISDIDDV 60
DB 1 MYLLVDVGNTHSVFSITEDGKTRFRWRLSTGVFQTEDE---LFSHLPLLGADAMREIKGI 57
QY 61 IISVVVPMHSLNFCIKYCKKOPLIYVPGIKTGL---NIKVDNPKQVQVADRIYNAVA 117
DB 58 GVASVVPQNTVIERPSQKYFHLPIWV--KAKNGCVKWNVK--NPSEVGADRIYNAVAF 113
QY 118 IEKYGAPSLVDGFTATTCFCAISEKGEYLGTTIAPGKISSEALFQSAKSLPRVELAKPG 177
DB 114 VKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFNMVHSLFRGTAKLPLVEVKPAD 172
QY 178 WTKICKSTVSAMOSGIYGVGLVDKIISIMKKELNCCDDVKVIATGGGLAKLIASETKSIDY 237
DB 173 FVVGKDTTEENIRLVGVNGSVYALEGLIGRIKEYVG--DLPVVLITGQSKIVKDMIKH-BI 229
QY 238 VDGFLTLEGL 247
DB 230 FDEDLTIKGV 239

Search completed: June 24, 2003, 22:16:17
Job time : 23.5189 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 10.3866 Seconds
(without alignments)
725.188 Million cell updates/sec

Title: US-09-813-453a-55
Perfect score: 1305
Sequence: 1 MLLVFDVGNMVLGIYKGD.....YVDGFLTEGLRIIYERNOE 256

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/1aa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101.5	7.8	2404	US-09-134-001C-3464	Sequence 3464, Ap
2	94	7.2	1529	US-09-134-001C-3945	Sequence 3945, Ap
3	93	7.1	381	US-09-134-001C-3003	Sequence 3003, Ap
4	92.5	7.1	330	US-09-134-001C-3582	Sequence 3582, Ap
5	89	6.8	529	US-09-504-358-22	Sequence 22, Appl
6	89	6.8	529	US-09-954-314-22	Sequence 22, Appl
7	88	6.7	514	US-09-134-001C-4347	Sequence 4347, Ap
8	87.5	6.7	546	US-09-068-047-2	Sequence 2, Appl
9	85.5	6.6	334	US-08-975-762-28	Sequence 28, Appl
10	85.5	6.6	334	US-08-821-324-28	Sequence 28, Appl
11	85.5	6.6	334	US-09-295-028-28	Sequence 28, Appl
12	85.5	6.6	334	US-09-106-582-28	Sequence 28, Appl
13	85.5	6.6	590	US-08-975-762-64	Sequence 64, Appl
14	85.5	6.6	590	US-09-295-028-64	Sequence 64, Appl
15	85.5	6.6	590	US-09-106-582-64	Sequence 64, Appl
16	84.5	6.5	731	US-08-911-364-1	Sequence 1, Appl
17	84.5	6.5	733	US-08-464-700-2	Sequence 2, Appl
18	84	6.4	553	US-08-663-566A-13	Sequence 13, Appl
19	84	6.4	553	US-08-484-575A-14	Sequence 14, Appl
20	84	6.4	553	US-08-023-610-13	Sequence 13, Appl
21	84	6.4	553	US-08-288-065A-13	Sequence 13, Appl
22	84	6.4	553	US-08-362-240A-13	Sequence 13, Appl
23	84	6.4	553	US-08-477-459-14	Sequence 14, Appl
24	84	6.4	553	US-08-479-869-14	Sequence 14, Appl
25	84	6.4	553	US-08-486-414-14	Sequence 14, Appl
26	84	6.4	553	US-08-804-372A-11	Sequence 11, Appl
27	84	6.4	553	PCT-US94-01826A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-09-134-001C-3464
; Sequence 3464, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3464
; LENGTH: 2404
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3464

Query Match	7.8%	Score 101.5	DB 4	Length 2404
Best Local Similarity	25.4%	Pred. No. 0.096	91	Indels 61
Matches	65	Conservative 39	Mismatches	91
Gaps	16			
QY	4	VFDVGNMVLGIYKGDKLNNYWRKTDREKTSDEYGLISNLFYDNVNSIDDDVITS	63	
DB	1681	VFDASATDFYCSLLNGYPLV---TATSVERTND---LLEKLISEQENTITAS-----IP	1698	
QY	64	SVPVNMHSLNFCIKYCKKOPLIYVPGIKTGLNTKYNPKOVGADRIYNAVAGIEKYGA	123	
DB	1699	LQVYVMHMF-----YIPK---VITGAPSTPAFVQHISKH---CDMYVNA-----YC-	1740	
QY	124	PSILVDFGTATFCAISEKGEYLGITAPGKISSEALF-QSASKLPVELAKPGMTICK	182	
DB	1741	PS-----ENTVITTSWIYKGDPAISTPIGKPLANVDIFIMSGGLGCVGI--PG-ELCI	1793	
QY	183	STVSAMSGIIVGYGLVDKIISIMKELNCDDV-----KVATGGLAKLIA-SEFK	233	
DB	1794	AGES-----LISGYLN-----RPLSAEKFINNPGPQLYRSGDLARLMPDGOIE	1839	
QY	234	SIDYVDGFLTEGLRI	249	
DB	1840	FLGRIDKQKVHGYRI	1855	

RESULT 2

US-09-134-001C-3945
; Sequence 3945, Application US/09134001C

```

; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3945
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3945

Query Match          7.28; Score 94; DB 4; Length 1529;
Best Local Similarity 26.99; Pred. No. 0.33;
Matches 39; Conservative 26; Mismatches 58; Indels 22; Gaps 5;

Qy 92 IKTGLNKKYDNPKQGVADRIVNAVAGTEKYGAPSILVDFGTATTCFAISEKGEYLGGTIA 151
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 665 IREGLMETSIIAQSGETREVVHVACILIGGANAV-VPYLAQRTIEQLTROGQ-LSGTVA 722

Qy 152 PGKISSEALFQSASKLPRLVELAKPGMTICKSTVMSQSGIIGYGVGLVDKLIISIMKEL 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 723 ENVATYTNVLSEGVIKY----MAKMGTI----STVQSYQAGIFEAVGLSNSVI----- 767

Qy 212 NCDDVKVIATGGLAKLIASETKSID 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 768 -----EKYFTGTSKLSGISIEQID 787

RESULT 3
US-09-134-001C-3003
; Sequence 3003, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3003
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3003

```

Qy	162	FQSASKLPRLVELA-----KPGMTICKSTVSAMQGIIYGVGL----	199
	:	: : :	
Dd	272	FETAGVVPAMKAVAQTIRGCTTTTGLPNPKDNFSPQVTLAAERTIKGSYVGSCVPD	331
	:	: : :	
Qy	200	--VDKIISIMKK--ELNCDDVKVIATGGGLAKLIASETSIDVV-DGF	241
	:	: : :	
Dd	332	RDIPIRFYNLYNOGRNLIDS-----LISEVTILDEINEGF	365

RESULT 4
US-09-134-001C-3582
; Sequence 3582, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3582
; LENGTH: 330
; TYPE: PRP
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3582

RESULT 5
US-09-504-358-22
: Sequence 22, Application US/09504358
: Patent No. 6365376
: GENERAL INFORMATION:
: APPLICANT: Rouviere, Pierre E.
: APPLICANT: Brostowicz, Patricia C.
: TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES
: FILE REFERENCE: BC1001 US NA
: CURRENT APPLICATION NUMBER: US/09/504,358
: CURRENT FILING DATE: 2000-02-15
: EARLIER APPLICATION NUMBER: 60/120,702

EARLIER FILING DATE: 1999-February-19
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Microsoft Office 97
SEQ ID NO 22
LENGTH: 529
TYPE: PRF
ORGANISM: Brevibacterium sp HCU
US-09-504-358-22

Query Match
Best Local Similarity 6.8%; Score 89; DB 4; Length 529;
Matches 51; Conservative 39; Mismatches 67; Indels 56; Gaps 14;
QY 4 VFDGNTNNVLGIYKG-----DKLVNWRKT-DR---EKTSD-----EYGI- 41
Db 283 MWDEGEFLWLNFGGLTDEAANTFYFNRKSVHVRKDPKTAEMLAPATPPHPGVK 342
QY 42 ---LISNLFDD---YDNNISDIDVLISSVVPNMHSLNFCIKYCKQPLIVGPGIKTG 95
Db 343 RPSLEQNTFDVYNQDNVLDISNATPITRVLPGVETPDG--VVECDVLVLATGFDNNSG 400
QY 96 ---LNIKYDNPQGVADRI---VNAVAGIEKYGAPSLVDFG--TATFCAISEKGEYL 146
Db 401 GINAIDIRAGG--QLLRDKWATGVDTYMGSLTHGPNLMFLYGPQSPSGFC-----NGTDF 454
QY 147 GGTIAPG-----IKISSEALFQSAKLPR 170
Db 455 GG--APGDWADFLWLKDNIGISREESTEEVER 485

RESULT 6

US-09-954-314-22
Sequence 22, Application US/09954314
Patent No. 6465224
GENERAL INFORMATION:
APPLICANT: Rouviere, Pierre E.
TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES
FILE REFERENCE: BC1001 US NA
CURRENT APPLICATION NUMBER: US/09/954,314
CURRENT FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: 60/120,702
PRIOR FILING DATE: 1999-February-19
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Microsoft Office 97
SEQ ID NO 22
LENGTH: 529
TYPE: PRF
ORGANISM: Brevibacterium sp HCU
US-09-954-314-22

Query Match
Best Local Similarity 6.8%; Score 89; DB 4; Length 529;
Matches 51; Conservative 39; Mismatches 67; Indels 56; Gaps 14;
QY 4 VFDGNTNNVLGIYKG-----DKLVNWRKT-DR---EKTSD-----EYGI- 41
Db 283 MWDEGEFLWLNFGGLTDEAANTFYFNRKSVHVRKDPKTAEMLAPATPPHPGVK 342
QY 42 ---LISNLFDD---YDNNISDIDVLISSVVPNMHSLNFCIKYCKQPLIVGPGIKTG 95
Db 343 RPSLEQNTFDVYNQDNVLDISNATPITRVLPGVETPDG--VVECDVLVLATGFDNNSG 400
QY 96 ---LNIKYDNPQGVADRI---VNAVAGIEKYGAPSLVDFG--TATFCAISEKGEYL 146
Db 401 GINAIDIRAGG--QLLRDKWATGVDTYMGSLTHGPNLMFLYGPQSPSGFC-----NGTDF 454
QY 147 GGTIAPG-----IKISSEALFQSAKLPR 170
Db 455 GG--APGDWADFLWLKDNIGISREESTEEVER 485

RESULT 7

US-09-134-001C-4347
Sequence 4347, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4347
LENGTH: 514
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4347

Query Match
Best Local Similarity 6.7%; Score 88; DB 4; Length 514;
Matches 63; Conservative 42; Mismatches 95; Indels 114; Gaps 15;
QY 40 GILISNLFYDYN--VNISDIDVLISSVVPNMHSLNFCIKYCKQPLIVGPGIKTGLN 97
Db 182 GMMNLEQDNDWKNKALELIGIERNQLPQVPTT-HVLTGMKKRYA-----TLMG 228
QY 98 IKYDNPQGVADRIYNAVAGIEKYGAPSLVDFGTA-----TTFCAISE 141
Db 229 IDEQTPVIVGASDGVLSNLGVNSYQGEVAVTIGTSGAIRTVINQPKTDEKGRIFCYILD 288
QY 142 KGEYL-----GGTIAPGK---ISSE-----ALFQSASKL-PRVE--LAK 175
Db 289 KDQYVIGGPNNGGVVLRWLRLDEILASEVETAKRLGVDYDVLTOIASRKVPGAEGLIFH 348
QY 176 P-----GMTIC---KSTVSAMOSGIIYGVGLVDRKIIIMKKELNC 213
Db 349 PYLAGERAPLNNADARGSFGLTLSHKKEHMTAALEGVLYNLYTVLALIEVM---NE 404
QY 214 DDVKVIATGGLAK-----LIASET-----KSIDVVDGFLTLE 245
Db 405 TPTTIKATGGFAKSEIRQWADIFDTDLIVPESVSSCLGACVLGMKALGEIDDFSVIK 464
QY 246 ---GLRIYKNOE 256
Db 465 DMVGTTHAEPNQE 478

RESULT 8

US-09-066-047-2
Sequence 2, Application US/09066047A
Patent No. 6306394
GENERAL INFORMATION:
APPLICANT: MURPHY, Cheryl
STOREY, James
BELTZ, Gerald A.
COUGHLIN, Richard T.
TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF USE OF GRANULOCYTIC ERLICHIA
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: HALE AND DORR LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: United States
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066.047A
FILING DATE: 24-Apr-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/044,869
FILING DATE: 25-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Superko, Colleen
REGISTRATION NUMBER: 39,850
REFERENCE/DOCKET NUMBER: 106,941.156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-066-047-2

Query Match 6.7%; Score 87.5; DB 4; Length 546;
Best Local Similarity 22.9%; Pred. No. 0.35;
Matches 66; Conservative 48; Mismatches 83; Indels 91; Gaps 16;
QY 4 VFDVGNMNLVGIYKGD-----KLVNYWRI-----KTDREKTSDEYG----- 40
DB 237 VLPVGVNRVVFALNEGEVSEMFVSVGVGHIMKVIKHEITKEDLEKLEKISSNIRROKA 296
QY 41 --ILISNLFYDYNV-----NISDIDVI---ISSVVPNV-MHSLNFCIKYCKKQPLVIG 89
DB 297 GELLVSNVKKANDMISRGASLNELKDMFGARISGLTDFMDHGLD----- 341
QY 90 PGIKTGLINKYDNPQVKGADRIVNA---VAGIEKYGAPSLVDGFTATTCALISEKGEYL 146
DB 342 ---KSG-NLVKDFPLQIG---INAFITLAFSSAVGKPSHLVNGDAYFGVLVTEVVP 393
QY 147 GGTIAPGIIKISSALFOSAKLPR-----VELAKPGMTICKSVSAMQGIYGVYGLVD 201
DB 394 PRTLEESRILTEE-WKSALRMKKIREFAVELR-----SKLQNGTELSVYNGVS 441
QY 202 --KIISIMKK-----ELNCDDVKVIATGGLAK-LIASETKSI 235
442 FRKNVTVKKSGDSTDNDKYPRLVDEIFAINGGVTKVEIDSETV 489

RESULT 9
US-08-975-762-28
Sequence 28, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,324
FILING DATE: 21-MAR-1997

TREATMEN

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 334 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-08-975-762-28

Query Match 6.6%; Score 85.5; DB 4; Length 334;
Best Local Similarity 22.9%; Pred. No. 0.27;
Matches 66; Conservative 44; Mismatches 87; Indels 91; Gaps 15;
QY 4 VFDVGNMNLVGIYKGD-----KLVNYWRI-----KTDREKTSDEYG----- 40
DB 80 VLPVGVNRVVFALNEGEVSEMFVSVGVGHIMKVIKHEITKEDLEKLEKISSNIRROKA 139
QY 41 --ILISNLFYDYNV-----NISDIDVI---ISSVVPNV-MHSLNFCIKYCKKQPLVIG 89
DB 140 GELLVSNVKKANDMISRGALLNELKDMFGARISGLTDFMDHGLD----- 184
QY 90 PGIKTGLINKYDNPQVKGADRIVNA---VAGIEKYGAPSLVDGFTATTCALISEKGEYL 146
DB 185 ---KSG-NLVKDFPLQIG---INAFITLAFSSAVGKPSHLVNGDAYFGVLVTEVVP 236
QY 147 GGTIAPGIIKISSALFOSAKLPR-----VELAKPGMTICKSVSAMQGIYGVYGLVD 201
DB 237 PRTLEESRILTEE-WKSALRMKKIREFAVELR-----SKLQNGTELSVYNGVS 284
QY 202 --KIISIMKK-----ELNCDDVKVIATGGLAK-LIASETKSI 235
DB 285 FRKNVTVKKSGDSTDNDKYPRLVDEIFAINGGVTKVEIDSETV 332

RESULT 10
US-08-821-324-28
Sequence 28, Application US/08821324
Patent No. 6231869
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,324
FILING DATE: 21-MAR-1997

THERAP

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 334 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Ehrlichia

US-08-821-324-28

Query Match 6.6%; Score 85.5; DB 4; Length 334;
Best Local Similarity 22.9%; Pred. No. 0.27;
Matches 66; Conservative 44; Mismatches 87; Indels 91; Gaps 15;

QY 4 VFDVGNNTNVLGIYKGD-----KLVNYWRI-----KTDREKTSDEYG----- 40
DB 80 VLPVGNRVNVPALNEGEVSEMFVSVGHIMKVIKHEITKEDLEKLEKISSNIRROKA 139
QY 41 --ILISNLFQYDNV-----NISDIDDVLISSVVPNV-MHSLNFCIKYCKKQPLIVG 89
DB 140 GELLVSNVKKANDMISRGALLNELKDMFGARISGVLTFNDFMHGLD----- 184
QY 90 PGKITGLNIKYDNPKQVQADRIANA---VAGIERKYGAPSIILVDFGTATTFCAISEKGEYL 146
DB 185 ---KSG-NLVKDFPLQLG---INAFITLAFSSAVGKPSHLVNSNGDAYEGLVLTVEVPPR 236
QY 147 GGTIAPGKISSEALFQASKLPR-----VELAKPGMTICKTSVAMQSGIYGVYGLVD 201
DB 237 PRTLEERSILTEE-WKSALRMKKIRFAVELR-----SKLQNGTSLSVVNGVS 284
QY 202 --KIISIMKK-----ELNCDDVKVIATGGLAK-LIASETKSI 235
DB 285 FKNVTVKSDGSTDNDKPYERLVDLFAINIGGVTKVEIDSETV 332

RESULT 11

US-09-295-028-28
Sequence 28, Application US/09295028
Patent No. 6277381

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION
FILE REFERENCE: 210121.439C4
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 28
LENGTH: 334
TYPE: PRT
ORGANISM: Ehrlichia sp.

US-09-295-028-28

Query Match 6.6%; Score 85.5; DB 4; Length 334;
Best Local Similarity 22.9%; Pred. No. 0.27;
Matches 66; Conservative 44; Mismatches 87; Indels 91; Gaps 15;

QY 4 VFDVGNNTNVLGIYKGD-----KLVNYWRI-----KTDREKTSDEYG----- 40
DB 80 VLPVGNRVNVPALNEGEVSEMFVSVGHIMKVIKHEITKEDLEKLEKISSNIRROKA 139

QY 41 --ILISNLFQYDNV-----NISDIDDVLISSVVPNV-MHSLNFCIKYCKKQPLIVG 89
DB 140 GELLVSNVKKANDMISRGALLNELKDMFGARISGVLTFNDFMHGLD----- 184
QY 90 PGKITGLNIKYDNPKQVQADRIANA---VAGIERKYGAPSIILVDFGTATTFCAISEKGEYL 146
DB 185 ---KSG-NLVKDFPLQLG---INAFITLAFSSAVGKPSHLVNSNGDAYEGLVLTVEVPPR 236
QY 147 GGTIAPGKISSEALFQASKLPR-----VELAKPGMTICKTSVAMQSGIYGVYGLVD 201
DB 237 PRTLEERSILTEE-WKSALRMKKIRFAVELR-----SKLQNGTSLSVVNGVS 284
QY 202 --KIISIMKK-----ELNCDDVKVIATGGLAK-LIASETKSI 235
DB 285 FKNVTVKSDGSTDNDKPYERLVDLFAINIGGVTKVEIDSETV 332

RESULT 12

US-09-106-582-28
Sequence 28, Application US/09106582
Patent No. 6306402

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

THERAP

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 334 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Ehrlichia

US-09-106-582-28

Query Match 6.6%; Score 85.5; DB 4; Length 334;
Best Local Similarity 22.9%; Pred. No. 0.27;
Matches 66; Conservative 44; Mismatches 87; Indels 91; Gaps 15;

QY 4 VFDVGNNTNVLGIYKGD-----KLVNYWRI-----KTDREKTSDEYG----- 40
DB 80 VLPVGNRVNVPALNEGEVSEMFVSVGHIMKVIKHEITKEDLEKLEKISSNIRROKA 139
QY 41 --ILISNLFQYDNV-----NISDIDDVLISSVVPNV-MHSLNFCIKYCKKQPLIVG 89
DB 140 GELLVSNVKKANDMISRGALLNELKDMFGARISGVLTFNDFMHGLD----- 184

Query Match	6.6%	Score 85.5;	DB 4;	Length 590;	
Best Local Similarity	22.9%	Pred. No. 0.68;			
Matches	66;	Conservative 44;	Mismatches 87;	Indels 91; Gaps 15;	
QY	4	VFDVGNTNMVLGYKGD----	KLNVKWRI-----	KTDREKTSDEYG-----	40
DB	336	VLPVGVRNVFALNEGEVSEMFPSVVGWHIMKVIKKEITKEDLEKKEISSNIRRQKA			395
QY	41	--TLISNLPDYDNV-----	NISDIDVLISSVPPNV--	MISLENFCIKYCKKQPLIIVG	89
DB	396	GELLSNVKKNAMISRGALLNELKDMFGARISGLVTFNDFMGLD-----			440
QY	90	PGIKTGLNIKYDNPKVQGDRIYNA----	VAGIEKYGAPSLYDFG	TATTFCAISEKGEYL	146
DB	441	---KSG-NLVKDPPLQLG----	INAFITLAFSSAVGKPSHLVNSGDAYFGLVTEVYVPPR		492
QY	147	GGTIAPIGKIKISSEALPQASAKLPR----	VELAKPCMTICKSTVPSAMQSGIIVGYGLVD		201
DB	493	PTLEERSRLTEE--WKSALRMKRIREFAVELR-----		SKLQNGTELSVNVGVS	540
QY	202	--KIISIMKK-----	ELNCDDDVKVIATGGGLAK--	LIASETKSI	235
DB	541	FKNVNVKKSDDGSTDNDSKYPERLVDIEPAINGVITKEVIDSESTV			588

```

RESULT 15
US-09-106--582-64
; Sequence 64, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 590 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-582-64

Query Match 6.6%; Score 85.5; DB 4; Length 590;
Best Local Similarity 22.9%; Pred. No. 0.68;
Matches 66; Conservative 44; Mismatches 87; Indels 91; Gaps 15;

Qy	4	VFDVGNMVLGIYKGD-----KLVNYWRI-----KTDREKTSDEYG-----	40
Db	336	VLPVGVNRVVFALNEGEVSEMFSGVGHIMKVIKHEITKEDLEKLEKISSNIRRQKA	395
Qy	41	--ILISNLFYDNY-----NISDIDDVISSVDPNV-MHSLNFCIKYCKKQPLIVG	89
Db	396	GELLVSNVKKANDMISRGNALLNELKDMFGARISGLTNFDMHGLD-----	440
Qy	90	PGIKTGLNINYPKQVGNDRVNA---VAGIEKYGAPSIYDFGTATTCATSEKGEYL	146
Db	441	---KSG-NLVKDFPLQLG---INAFITLAFSSAVGKPSHLVSNGLDAYFGVLVTEVVP	492
Qy	147	GGTAPGKISSSEALFOSASKLPR-----VELAKPGMTICKSVSAMQSGIIVGYVGLVD	201
Db	493	PRTLEERSILTEP-WKSALRMKKIREFAVELR-----SKLQNGTELSVNVGS	540
Qy	202	--KIISIMKK-----ELNCDDVKVIATGGLAK-LIASETKSI	235
Db	541	EKKNVYKSDGSTDNDNSKYPERLVDFEFAINIGGVTKVIDSESEIV	588

Search completed: June 24, 2003, 22:10:46
Job time : 12.3866 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.2271 Seconds

(without alignments)
1947.059 Million cell updates/sec

Title: US-09-813-453A-55

Perfect score: 1305

Sequence: 1 MLLVFDVGNVNLGIYKGD.....YVDGFLTEGLRIIYKQNE 256

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 41779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1305	100.0	256	9	US-09-813-453A-55
2	817.5	62.6	258	9	US-09-813-453A-49
3	814.5	62.4	254	9	US-09-813-453A-47
4	802.5	61.5	262	9	US-09-813-453A-45
5	794.5	60.9	258	9	US-09-813-453A-2
6	785.5	60.2	255	9	US-09-813-453A-7
7	695	53.3	255	9	US-09-813-453A-3
8	672	51.5	233	9	US-09-813-453A-17
9	643.5	49.3	265	9	US-09-813-453A-4
10	523.5	40.1	260	9	US-09-813-453A-51
11	513	39.3	272	9	US-09-813-453A-5
12	509	39.0	272	9	US-09-712-363-276
13	496.5	38.0	258	9	US-09-813-453A-6
14	489	37.5	219	9	US-09-813-453A-57
15	373.5	27.9	262	9	US-09-813-453A-8
16	354	27.1	273	9	US-09-813-453A-10
17	321	24.6	246	9	US-09-813-453A-9
18	315.5	24.2	212	9	US-09-813-453A-59
19	282	21.6	262	9	US-09-813-453A-11

20	254.5	19.5	257	9	US-09-813-453A-53	Sequence 53, Appl
21	201.5	15.4	244	9	US-09-813-453A-41	Sequence 41, Appl
22	145	11.1	241	9	US-09-813-453A-61	Sequence 61, Appl
23	144.5	11.1	257	9	US-09-813-453A-13	Sequence 13, Appl
24	140.5	10.8	592	9	US-09-813-453A-22	Sequence 22, Appl
25	140.5	10.8	592	9	US-09-813-453A-39	Sequence 39, Appl
26	138.5	10.6	460	9	US-09-813-453A-12	Sequence 12, Appl
27	133	10.2	229	9	US-09-813-453A-12	Sequence 12, Appl
28	131	10.0	249	9	US-09-813-453A-61	Sequence 61, Appl
29	127.5	9.8	267	9	US-09-813-453A-15	Sequence 15, Appl
30	127	9.7	249	9	US-09-813-453A-70	Sequence 70, Appl
31	120	9.2	242	9	US-09-813-453A-65	Sequence 65, Appl
32	119.5	9.2	248	9	US-09-813-453A-20	Sequence 20, Appl
33	102.5	7.9	1159	10	US-09-815-242-11089	Sequence 11089, A
34	101	7.7	223	9	US-09-895-913A-74	Sequence 74, Appl
35	101	7.7	223	9	US-09-813-453A-14	Sequence 14, Appl
36	101	7.7	223	9	US-09-813-453A-67	Sequence 67, Appl
37	89	6.8	529	10	US-09-954-314-22	Sequence 22, Appl
38	87.5	6.7	1160	10	US-09-815-242-13913	Sequence 13913, A
39	86.5	6.6	331	9	US-09-738-626-4960	Sequence 4960, Ap
40	86.5	6.6	1173	10	US-09-815-242-11935	Sequence 11935, A
41	85.5	6.6	334	10	US-09-159-469-28	Sequence 28, Appl
42	85.5	6.6	334	10	US-09-798-042-28	Sequence 28, Appl
43	85.5	6.6	590	10	US-09-159-469-64	Sequence 64, Appl
44	85.5	6.6	590	10	US-09-798-042-64	Sequence 64, Appl
45	84.5	6.5	1160	10	US-09-815-242-10047	Sequence 10047, A

ALIGNMENTS

RESULT 1
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-09-813-453A-55

Query Match	100.0%	Score 1305;	DB 9;	Length 256;
Best Local Similarity	100.0%	Pred. No. 3e-116;		
Matches 256;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLLVFDVGNVNLGIYKDKLVNYWRIKTDREKTSDEYGLISNLFYDNNVNISDIDDV	60	
Db	1	MLLVFDVGNVNLGIYKDKLVNYWRIKTDREKTSDEYGLISNLFYDNNVNISDIDDV	60	
QY	61	IISSVVPNMVHLENFCIKYCKKQPLIVGPGIKTGLNIKYDNPQKVGADRIYNAVAGIEK	120	
Db	61	IISSVVPNMVHLENFCIKYCKKQPLIVGPGIKTGLNIKYDNPQKVGADRIYNAVAGIEK	120	
QY	121	YGAPSLIVDFGATTFCAISEKGEYLGTTIAPGKISSSEALFQSAKSLPRVELAKPGMTI	180	
Db	121	YGAPSLIVDFGATTFCAISEKGEYLGTTIAPGKISSSEALFQSAKSLPRVELAKPGMTI	180	
QY	181	CKSTVSAMSGIIYGVGLVDKIIISIMKKELCDDVKVIATGGLAKLIASETKSIDYDVG	240	
Db	181	CKSTVSAMSGIIYGVGLVDKIIISIMKKELCDDVKVIATGGLAKLIASETKSIDYDVG	240	

Db. 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
Qy 241 FLTLEGLRIIYKNE 256
Db 241 FLTLEGLRIIYKNE 256

RESULT 2

US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: CGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

Query Match 62.6%; Score 817.5; DB 9; Length 258;

Best Local Similarity 62.9%; Pred. No. 7.1e-70;

Matches 161; Conservative 40; Mismatches 54; Indels 1; Gaps 1;

Qy 1 MLVFDVGNNTMVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFDDYDNNISDIDDV 60
Db 1 MIFLDVGNNTVLGVYQDETLLVHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
Qy 61 IISVVPNVHSLNFCIKYCKKQPLIVGPGIKTGLNIIKYNPNKQVADRVNNAVAGIEK 120
Db 61 IISVVPNVHSLNFCIKYCKKQPLIVGPGIKTGLNIIKYNPNKQVADRVNNAVAGIEK 120
Qy 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Db 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Qy 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Db 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Qy 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
Db 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
Qy 241 FLTLEGLRIIYKNE 256
Db 241 FLTLEGLRIIYKNE 256

RESULT 3

US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: CGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 47

; LENGTH: 254

; TYPE: PRT

; ORGANISM: Bacillus halodurans

US-09-813-453A-47

Query Match 62.4%; Score 814.5; DB 9; Length 254;

Best Local Similarity 60.2%; Pred. No. 1.3e-69;

Matches 153; Conservative 45; Mismatches 55; Indels 1; Gaps 1;

Qy 1 MLVFDVGNNTMVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFDDYDNNISDIDDV 60
Db 1 MLVFDVGNNTVLGVYQDETLLVHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
Qy 61 IISVVPNVHSLNFCIKYCKKQPLIVGPGIKTGLNIIKYNPNKQVADRVNNAVAGIEK 120
Db 61 IISVVPNVHSLNFCIKYCKKQPLIVGPGIKTGLNIIKYNPNKQVADRVNNAVAGIEK 120
Qy 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Db 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Qy 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
Db 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
Qy 241 FLTLEGLRIIYKNE 254
Db 241 FLTLEGLRIIYKNE 254

RESULT 4

US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: CGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match 61.5%; Score 802.5; DB 9; Length 262;

Best Local Similarity 59.1%; Pred. No. 1.9e-68;

Matches 150; Conservative 49; Mismatches 54; Indels 1; Gaps 1;

Qy 1 MLVFDVGNNTMVLGIYKDGKLVNRYRIKTDREKTSDEYGLISNLFDDYDNNISDIDDV 60
Db 1 MIFLDVGNNTVLGVYQDETLLVHWRLATSRQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60
Qy 61 IISVVPNVHSLNFCIKYCKKQPLIVGPGIKTGLNIIKYNPNKQVADRVNNAVAGIEK 120
Db 61 IISVVPNVHSLNFCIKYCKKQPLIVGPGIKTGLNIIKYNPNKQVADRVNNAVAGIEK 120
Qy 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Db 121 YGAPSLVDGTTATTCATSEKGEYLGTTIAPGIIKISSALFOSASKLPRVELAKPGMTI 180
Qy 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240
Db 181 CKSTVSAMQSGIYGVGLVDKIIISIMKKELNCDVVKVIATGGLAKLIASETSIDYVDG 240

Db 181 GNTVSAMQSGILYGVQVEGIVKRMKEAK-QBPKVIATGGLAKLISESNVIDVDP 239
QY 241 FLTLEGLRIIYEKN 254
Db 240 FLTLAGLYMYERN 253

RESULT 5
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 60.9%; Score 794.5; DB 9; Length 258;
Best Local Similarity 60.8%; Pred. No. 1.1e-67;
Matches 155; Conservative 40; Mismatches 59; Indels 1; Gaps 1;
QY 1 MLLVFDVGNNTNVLGIYKDGKLVNYWRIKTDREKTSDEYIGLISNLFYDYNVNISDIDV 60
Db 1 LLLVIDVGNNTNVLGVYHDKLEYHWRIETSHKTEDEFGMTLSLFDHSGLMFFQIDGI 60
QY 61 IISVVVPNMVMSLENFCIKYCKKQPLIVGPGIKTGLNPKYQVADRIYNAVAGIEK 120
Db 61 IISVVVPIMFALEMCYKTHIEPQIVGPGMKTGLNPKYQVADRIYNAVAAIHL 120
QY 121 YCAPSILVDFGTATTCFCAISEKGEVYLGTTIAPGKISSEALFQSKLPVELAKPGMTI 180
Db 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAPIITISTEALYSRAAKLPRIETRPDNI 180
QY 181 CKSTVSAMQSGIYGVYGLVDKIIISIMKELNCDVVKVIATGGLAKLIASETKSIDYVDG 240
Db 181 GNTVSAMQSGILYGVQVEGIVKRMKQAK-QDLKVIATGGLAPLIANESDCIDVDP 239
QY 241 FLTLEGLRIIYEKN 255
Db 240 FLTLAGLYMYERN 254

RESULT 6
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 60.2%; Score 785.5; DB 9; Length 255;
Best Local Similarity 60.9%; Pred. No. 7.7e-67;
Matches 156; Conservative 44; Mismatches 55; Indels 1; Gaps 1;
QY 1 MLLVFDVGNNTNVLGIYKDGKLVNYWRIKTDREKTSDEYIGLISNLFYDYNVNISDIDV 60
Db 1 MLLVIDVGNNTNVLGIYDGERLVRDWRVSTDKARTDEYIGLISNLFRLAGLGLDIRAV 60
QY 61 IISVVVPNMVMSLENFCIKYCKKQPLIVGPGIKTGLNPKYQVADRIYNAVAGIEK 120
Db 61 IISVVVPPLTGVLERLSLGYFGMRPLVGVPGIKTGMPIQYDNPREVADRIYNAVAGYEK 120
QY 121 YCAPSILVDFGTATTCFCAISEKGEVYLGTTIAPGKISSEALFQSKLPVELAKPGMTI 180
Db 121 YRTSLIVDFGTATTFDYVNRAGEYCGGAIAPELVISTEALFQSKLPVRDIIRPSAI 180
QY 181 CKSTVSAMQSGIYGVYGLVDKIIISIMKELNCDVVKVIATGGLAKLIASETKSIDYVDG 240
Db 181 ARNTVNSMQAGIYGVYGLVDVIVTRMAE-SKDAPRVATGGLASLIAPESKTIIEAVEE 239
QY 241 FLTLEGLRIIYEKN 256
Db 240 YLTLEGLRIIYERNRE 255

RESULT 7
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

Query Match 53.3%; Score 695; DB 9; Length 250;
Best Local Similarity 59.0%; Pred. No. 3e-58;
Matches 138; Conservative 37; Mismatches 51; Indels 8; Gaps 2;
QY 1 MLLVFDVGNNTNVLGIYKDGKLVNYWRIKTDREKTSDEYIGLISNLFYDYNVNISDIDV 60
Db 18 VILVLDVGNNTNVLGIYNDTKLTAERLSTDLVRSADYEGIQVWNLFDQDKLDPTLVEGV 77
QY 61 IISVVVPNMVMSLENFCIKYCKKQPLIVGPGIKTGLNPKYQVADRIYNAVAGIEK 120
Db 78 IISVVVPNMVMSLENFCIKYCKKQPLIVGPGIKTGLNPKYQVADRIYNAVAAHEI 137
QY 121 YCAPSILVDFGTATTCFCAISEKGEVYLGTTIAPGKISSEALFQSKLPVELAKPGMTI 180
Db 138 YKRSLLIIDFGTATTCFCAVRENGDYLGGAICPGIKVSSSEALFEKAKLPVELIKPAYAI 197


```

RESULT 12
US-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712.363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593

```

```

Query Match      38.0%; Score 496.5; DB 9; Length 258;
Best Local Similarity 41.4%; Pred. No. 2.4e-39;
Matches 106; Conservative 49; Mismatches 98; Indels 3; Gaps 3;

QY 1 MLLVFDVGNVMVLGIYKGDKLNVYRIKTDREKTSDEYTGILISNLFEDYDNNVISIDDDV 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MLCIDCGNTVNFVSGVDGDFDAATWRIATDTHRTADEYFVWLNTLMQLQGLQ-GRISA 59
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 29.2042 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453a-49

Perfect score: 1321

Sequence: 1 MIFVLDVGNNTVTLGVYDGD.....PFLTLGLKLLYKNTKKG 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1321	100.0	258	AAU91172	Pantothenate kinas
2	1057.5	80.1	262	AAU91170	Pantothenate kinas
3	1046	79.2	258	AAU01243	B. subtilis novel
4	1046	79.2	258	AAU91149	Bacillus subtilis
5	975	73.8	254	AAU91171	Pantothenate kinas
6	900	68.1	233	AAU91163	Pantothenate kinas
7	855.5	64.8	259	AB47661	Listeria monocytog
8	834	63.1	255	AAU91154	Geobacter sulfurre
9	817.5	61.9	256	AAU91175	Pantothenate kinas
10	663	50.2	265	AAU91151	Streptomyces coeli

11	629.5	47.7	250	23	AAU91150	Clostridium acetob
12	575	43.5	260	23	AAU91173	Pantothenate kinas
13	542	41.0	258	23	AAU91153	Rhodobacter capsul
14	513	38.8	219	23	AAU91176	Pantothenate kinas
15	479.5	36.3	272	23	AAU91152	Mycobacterium tube
16	475.5	36.0	272	22	AAU91155	Mycobacterium radiop
17	415.5	31.5	262	23	AAU91155	Delnocooccus radiop
18	375.5	28.4	246	23	AAU91156	Thermotoga maritim
19	352.5	26.7	212	23	AAU91177	Pantothenate kinas
20	342.5	25.9	273	23	AAU91157	Treponema pallidum
21	305	23.1	257	23	AAU91174	Pantothenate kinas
22	256.5	19.4	262	23	AAU91158	Borrelia burgdorfe
23	210.5	15.9	244	23	AAU91168	Pantothenate kinas
24	178.5	13.5	241	23	AAU91179	Pantothenate kinas
25	157	11.9	249	23	AAU91182	Pantothenate kinas
26	150	11.4	257	23	AAU91160	Synechocystis pant
27	148.5	11.2	229	23	AAU91159	Aquifex aeolicus p
28	146	11.1	389	21	AAU91159	Neisseria meningit
29	146	11.1	455	21	AAU91159	Neisseria meningit
30	146	11.1	592	20	AAU91159	Neisseria meningit
31	146	11.1	592	20	AAU91159	Neisseria meningit
32	146	11.1	592	21	AAU91159	Neisseria meningit
33	146	11.1	592	21	AAU91159	Neisseria meningit
34	146	11.1	592	23	AAU91166	Pantothenate kinas
35	146	11.1	592	23	AAU91169	Pantothenate kinas
36	145	11.0	455	20	AAU91169	Neisseria gonorrhoe
37	145	11.0	455	21	AAU91169	Neisseria gonorrhoe
38	145	11.0	460	23	AAU91167	Pantothenate kinas
39	145	11.0	592	20	AAU91167	Neisseria gonorrhoe
40	145	11.0	592	21	AAU91167	Neisseria gonorrhoe
41	140.5	10.6	267	23	AAU91162	Bordetella pertussis
42	139	10.5	249	23	AAU91178	Pantothenate kinas
43	123.5	9.3	248	23	AAU91164	Pantothenate kinas
44	115	8.7	242	23	AAU91180	Pantothenate kinas
45	105.5	8.0	189	20	AAU91164	Neisseria meningit

ALIGNMENTS

RESULT 1

AAU91172
ID AAU91172 standard; Protein; 258 AA.

XX AAU91172;

XX 05-JUN-2002 (first entry)

DE Pantothenate kinase (Coax) #10.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Bacillus stearothermophilus.

XX WO200216601-A2.

PD 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX N-PSDE; ABK54193.

PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein -
PS Claim 10; Page 101-102; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX SQ Sequence 258 AA;
Query Match 100.0%; Score 1321; DB 23; Length 258;
Best Local Similarity 100.0%; Pred. No. 4.7e-141;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MIFVLDVGNNTVGLVYDGDDELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
1 MIFVLDVGNNTVGLVYDGDDELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
61 IISVVPPIMFALERCLYFHKPLIVGPGIKTGLDIKYNPREVGADRIYNVAGIHL 120
61 IISVVPPIMFALERCLYFHKPLIVGPGIKTGLDIKYNPREVGADRIYNVAGIHL 120
121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAALPRIETARPDDII 180
121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAALPRIETARPDDII 180
181 GNTVSAMQAGILYGVGVGEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDDPF 240
181 GNTVSAMQAGILYGVGVGEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDDPF 240
241 LTLTGLKLLYERKNTKKG 258
241 LTLTGLKLLYERKNTKKG 258
RESULT 2
AAU91170
ID AAU91170 standard; Protein; 262 AA.
XX AAU91170;
AC AAU91170;
DT 05-JUN-2002 (first entry)
Pantothenate kinase (Coax) #8.
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Bacillus anthracis.
OS WO200216601-A2.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI: 2002-269358/31.
XX N-PSDB; ABK54191.
XX

PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX Claim 8; Page 98-99; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX SQ Sequence 262 AA;
Query Match 80.1%; Score 1057.5; DB 23; Length 262;
Best Local Similarity 77.0%; Pred. No. 3.4e-111;
Matches 201; Conservative 26; Mismatches 31; Indels 3; Gaps 1;
QY 1 MIFVLDVGNNTVGLVYDGDDELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
DB 1 MIFVLDVGNNTVGLVYDGDDELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
QY 61 IISVVPPIMFALERCLYFHKPLIVGPGIKTGLDIKYNPREVGADRIYNVAGIHL 120
DB 61 IISVVPPIMFALERCLYFHKPLIVGPGIKTGLDIKYNPREVGADRIYNVAGIHL 120
QY 121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAALPRIETARPDDII 180
DB 121 YGSPLIIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAALPRIETARPDDII 180
QY 181 GNTVSAMQAGILYGVGVGEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDDPF 240
DB 181 GNTVSAMQAGILYGVGVGEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDDPF 240
QY 241 LTLTGLKLLYERKNT---EKKG 258
DB 241 LTLTGLKLLYERKNT---EKKG 258
RESULT 3
AAU01243
ID AAU01243 standard; Protein; 258 AA.
XX AAU01243;
AC AAU01243;
DT 18-JUL-2001 (first entry)
XX B. subtilis novel pantothenate kinase encoded by the gene coax.
DE Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;
KW nutritional supplement; panto-compound; pantoate.
XX Bacillus subtilis.
OS WO200121772-A2.
XX WO200121772-A2.
XX 29-MAR-2001.
XX 21-SEP-2000; 2000WO-US25993.
XX 21-SEP-1999; 99US-0400494.
PR 07-JUN-2000; 2000US-0210072.
PR 28-JUL-2000; 2000US-0221836.
PR 24-AUG-2000; 2000US-0227860.
XX (OMNI-) OMNIGENE BIOPRODUCTS.
XX Yocum RR, Patterson TA, Hermann T, Pero JG;
PI

[illegible]

PR 24-AUG-2000; 2000US-227860P.
 XX 20-MAR-2001; 2001US-0813453.
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.
 DR N-PSDB; ABK54192.
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 XX Claim 10; Page 100; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX SQ Sequence 254 AA;

Query Match 73.8%; Score 975; DB 23; Length 254;
 Best Local Similarity 70.8%; Pred. No. 7.2e-102;
 Matches 179; Conservative 36; Mismatches 38; Indels 0; Gaps 0;
 QY 1 MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
 DB 1 MILVIDGNTNTVLGVYQDETLLVHWRLATSRKTEDEYATVRSFLDHAGLQFQIDGI 60
 QY 61 IISVVPPIMFALERMCLYFHKPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHL 120
 DB 61 VISSVPPPMFSLQMKCKYFHVPTMIGPGIKTGLNKNYDNPKEVGADRIVNAVAGIEL 120
 QY 121 YGSPLLIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAAKLPRIETARPDII 180
 DB 121 YGPAIVDFGTATTCYINEKKQYAGVIAPGIMISTEALYHRASKLPRIETAKPKQV 180
 QY 181 GKNTVSAMQAGILYGVGVQVEGIVSRMAKSKIPPVKIATGGIAPLIASESDIIDVDPF 240
 DB 181 GTWIDSMQSGIFYGVYQVGVKMKQAASEPKVIATGGLAKLIGTSETIDVDSF 240
 DB 241 LTLTGKLLYEKN 253
 DB 241 LTLKGLQLYKKN 253

RESULT 6
 AAU91163
 ID AAU91163 standard; Protein; 233 AA.
 XX
 AC AAU91163;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Pantothenate kinase (Coax) #1.
 XX
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX
 OS Bacillus subtilis.
 XX
 PN WO200216601-A2.
 XX
 PD 28-FEB-2002.
 XX

PF 24-AUG-2001; 2001WO-US26531.
 XX
 PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 XX
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX Yocum RR, Patterson TA;
 XX WPI; 2002-269358/31.
 DR N-PSDB; ABK54169.
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 XX
 XX Disclosure; Page 81-82; 128pp; English.
 XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 XX SQ Sequence 233 AA;
 Query Match 68.1%; Score 900; DB 23; Length 233;
 Best Local Similarity 78.2%; Pred. No. 2e-93;
 Matches 169; Conservative 27; Mismatches 20; Indels 0; Gaps 0;
 QY 1 MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
 DB 1 MLVIDGNTNTVLGVYHDGKLEYHWRIETSRKTEDEFGMLRSLDFHSLMFEQIDGI 60
 QY 61 IISVVPPIMFALERMCLYFHKPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHL 120
 DB 61 IISVVPPIMFALERMCLYFHKPLIVGPGIKTGLDIKYNPREVGADRIVNAVAGIHL 120
 QY 121 YGSPLLIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAAKLPRIETARPDII 180
 DB 121 YGNPLIVDFGTATTCYIDENKQYMGGAIAFGITISTEALYSRAAKLPRIETRPDII 180
 QY 181 GKNTVSAMQAGILYGVGVQVEGIVSRMAKSKIPPX 216
 DB 181 GKNTVSAMQSGILFGYGVQVEGIVSRMAKQKQDP 216

RESULT 7
 ABB47661
 ID ABB47661 standard; Protein; 259 AA.
 XX
 AC ABB47661;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Listeria monocytogenes protein #365.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX

```
PR 11-APR-2000; 2000FR-0004629.
XX (INSP) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fshih H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablo B, Wehland U, Kaerst U, Entian K, Hauf J;
XX Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides
XX Claim 6; SEQ ID No 366; 192pp; French.
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGB-e (see ABA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccine compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 259 AA;
Query Match 64.8%; Score 855.5; DB 23; Length 259;
Best Local Similarity 62.9%; Pred. No. 2.6e-88;
Matches 163; Conservative 41; Mismatches 52; Indels 3; Gaps 1;
QY 1 MIFVLDVGNNTVILGVYDDELKHHWRITSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
DB 1 MLVIDVGNNTCTGVYKQKLLKHHWRITDRHRTSDELGMTVLNFFSYANLTSPDIQGI 60
QY 61 IISVVPPIPMFALERMCKLYPHIKPLIVGPGIKTGKLDIKYDNPREGADRIYNVAGIHL 120
DB 61 IISVVPPIPMAMETCMVRYNIRPLIVGPGIKTGKLNKUNDPREGSDRIYNVAASEE 120
QY 121 YGSPLIIVDFGTATTCYCYINEHKOYMGGAIAFGIMISTEALFARAAPKLPRIEIRPDDII 180
DB 121 YGTPVIVVDFGTATTCYCYIDESGVYQGAIAFGIMISTEALYNRAAKLPRIEIRPDDII 180
QY 181 GKNTVSAMQAGILYGVGVGVEGIVSRMKAISKIPPKVATATGLAPLIAESDIDVDVDF 240
DB 181 GKSTVSSMQAGIFYGVGCGGCIATMCKQSNASPVVATKGLARMITEKSAVDILDVDF 240
QY 241 LTLTGLKLLYEKN---TEK 256
DB 241 LTLKGLLELYRNKPTTEK 259
RESULT 8
AAU91154
ID AAU91154 standard; Protein; 255 AA.
XX AC AAU91154;
XX
```

```
DT 05-JUN-2002 (first entry)
XX Geobacter sulfurreducens pantothenate kinase Coax.
DE
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Geobacter sulfurreducens.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein
XX Claim 10; Page 72-73; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX SQ Sequence 255 AA;
Query Match 63.1%; Score 834; DB 23; Length 255;
Best Local Similarity 62.7%; Pred. No. 6.9e-86;
Matches 160; Conservative 40; Mismatches 55; Indels 0; Gaps 0;
QY 1 MIFVLDVGNNTVILGVYDDELKHHWRITSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
DB 1 MLVIDVGNNTVILGIYDGERLVRDVRVSTKARTTDEYGLINELFRLAGLGLQIRAV 60
QY 61 IISVVPPIPMFALERMCKLYPHIKPLIVGPGIKTGKLDIKYDNPREGADRIYNVAGIHL 120
DB 61 IISVVPPIPLTGLYERLSGLYFGMRPLVVGPGIKTGMPIQYDNPREGADRIYNVAGYEK 120
QY 121 YGSPLIIVDFGTATTCYCYINEHKOYMGGAIAFGIMISTEALFARAAPKLPRIEIRPDDII 180
DB 121 YGTPVIVVDFGTATTCYCYIDVNRKGEYCGGAIARGLVISTEALFQRAKLPRIEIRPDDII 180
QY 181 GKNTVSAMQAGILYGVGVGVEGIVSRMKAISKIPPKVATATGLAPLIAESDIDVDVDF 240
DB 181 ARNTVNSMQAGIYGVGLVDEIVTRMKAESKADPRVATGLASLIAESKTIKTIKTIKTIK 240
QY 241 LTLTGLKLLYEKNTE 255
DB 241 LTLGLRILYERNRE 255
RESULT 9
AAU91175
ID AAU91175 standard; Protein; 256 AA.
XX AC AAU91175;
XX
```

```
XX 05-JUN-2002 (first entry)
XX Pantothenate kinase (Coax) #13.
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Clostridium difficile.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX N-PSDB; ABK54196.
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein -
XX Claim 6; Page 105; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX Sequence 256 AA;
Query Match 61.9%; Score 817.5; DB 23; Length 256;
Best Local Similarity 62.9%; Pred. No. 5.1e-84;
Matches 161; Conservative 40; Mismatches 54; Indels 1; Gaps 1;
1 MIFVLDVGNNTVLGVYDGDGLKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
1 MLLVFDVGNNTVMVLGIYKGDGLVYWRKTRDKTSDEYGLISNLFYDYNVISDIDV 60
61 IISVVPPIPMFALERCLYFKHKLPLVPGIKTGLDIKYNPREVGADRIYNVAGIHL 120
1 IISVVPNVNHSLENCIYKQKPLVPGIKTGLNIRKYNPREVGADRIYNVAGIEK 120
121 YGSPILIVDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAAKLPRIETARPDDII 180
121 YGAPSLVDFGTATTCALISEKGEYLGGTAPGIIKISSALFQSAKLPRVELAKPGMTI 180
181 GNTVSAMQAGILYGVGVGVEGIVSRMKAKSKIPP- KVIATGGLAPLIASESDIIDVDP 239
181 CKSTVSAMQSGIYGVGVGVLDVKIISIMKKNCLNCDVVKVIATGGLAKLIASETSIDYVDG 240
240 FLTTLTGKLLYEKNT 255
241 FLTLEGLRIYKNOE 256
RESULT 11
AAU91151
ID AAU91151 standard; Protein; 265 AA.
```

```
XX AAU91151;
XX 05-JUN-2002 (first entry)
XX Streptomyces coelicolor pantothenate kinase Coax.
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Streptomyces coelicolor.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein -
XX Claim 10; Page 69-70; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX Sequence 265 AA;
Query Match 50.2%; Score 663; DB 23; Length 265;
Best Local Similarity 48.9%; Pred. No. 1.7e-66;
Matches 129; Conservative 58; Mismatches 69; Indels 8; Gaps 5;
1 MIFVLDVGNNTVLGVYDGDGLKHHWRIETSRKTEDEYGMKIKALLN-H--VGLQFSD- 56
1 MLLTIDVGNTHVVLGLFDGEDIVEHWRISTDSRRTRADELAVLLQGLMGHPLGLDELGDG 60
57 IRGIISSVVPPIPMFALERCLYFKHKLPLVPGIKTGLDIKYNPREVGADRIYNVAV 115
61 IDGATCATVSPSLHRELVTRYRGDPAVLVEPGVTKGTPILTDHPKEVGADRIYNVAV 120
116 AGIHLVGSPLIIVDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAAKLPRIETAR 175
121 AAVELYGGPAIVVDVGTATTFDAVSARGEIVTAPGIEISVEALGVKGAQLRKIEVAR 180
176 PDDITGKNTVSAMQAGILYGVGVGVEGIVSRMKAKSKIPP---KVIATGGLAPLIASESD 232
181 PRSVGKNTVEAMQSGIYGVGVGVLDVKIISIMKKNCLNCDVVKVIATGGLAPMWLGESS 240
233 IIDVVDVDFLTTLTGKLLYEKNT 256
241 VIDEHEPWLTLMLGLRLVYERNVSR 264
RESULT 11
AAU91150
```


AAU91150 standard; Protein: 250 AA.
AAU91150;
05-JUN-2002 (first entry)
Clostridium acetobutylicum pantothenate kinase Coax.
Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.
Clostridium acetobutylicum.
WO200216601-A2.
28-FEB-2002.
24-AUG-2001; 2001WO-US26531.
24-AUG-2000; 2000US-227860P.
20-MAR-2001; 2001US-0813453.
(OMNI-) OMNIGENE BIOPRODUCTS INC.
Yocum RR, Patterson TA;
WPI; 2002-269358/31.
Identifying potential antibiotic or antimicrobial agent, comprises
contacting composition comprising pantothenate kinase (Coax) protein
with test compound and identifying inhibitor of the Coax protein -
Claim 10; Page 68-69; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic
kinase (Coax) protein with a test compound, and determining the ability
of the test compound to inhibit the activity of the Coax protein, an
essential enzyme for the production of coenzyme A. Coax protein is a
valuable target for identifying bactericidal compounds. Coax modulating
agents can be used in an infectious animal model to determine the
efficacy, toxicity, or side effects of treatment with such an agent. This
is the amino acid sequence of a pantothenate kinase (Coax) protein
described in the invention.
Sequence 250 AA;
Query Match 47.7%; Score 629.5; DB 23; Length 250;
Best Local Similarity 53.8%; Pred. No. 9.8e-63;
Matches 126; Conservative 40; Mismatches 59; Indels 9; Gaps 2;
QY 1 MIFVLDVGNNTVILGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
DB 18 VILVDGNTNIVLGIYNDTALTAERLUSTDVLRSADYEGIQVNNLFQODKLDPTLVEGV 77
QY 61 IISVVPPIMPALERMCKLYPHIKPLIVGPGIKTGLDIKYDNPREGVADRVNVAAGIHL 120
DB 78 IISVVPPNIMYSLEHMIRKFKINPLVVGPGIKTGINKIDNPKEVGADRVNVAAGHEI 137
QY 121 YGSLPIIVDFGTATTTCYVINEHKQYMGGAIAPIGIMISTEALFARAALPRTIARPDII 180
DB 138 YKRSIIIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAALPRVELIKPAVAI 197
QY 181 GKNTVSAMOGILYGVGVGVEGIVSRMKAKSKIP-----PKVIATGGGLAPLI 227
DB 198 CKNTISSQSGIVYRLRQVKYLF--KLKENLPDGRTRTSLVATGGGLAKLI 249
RESULT 12
AAU91173
ID AAU91173 standard; Protein: 260 AA.
XX
AC AAU91173;

05-JUN-2002 (first entry)
Pantothenate kinase (Coax) #11.
Pantothenate kinase; Coax; antibiotic; antimicrobial;
pantothenate kinase modulator; coenzyme A; bactericidal compound.
Caulobacter crescentus.
WO200216601-A2.
28-FEB-2002.
24-AUG-2001; 2001WO-US26531.
24-AUG-2000; 2000US-227860P.
20-MAR-2001; 2001US-0813453.
(OMNI-) OMNIGENE BIOPRODUCTS INC.
Yocum RR, Patterson TA;
WPI; 2002-269358/31.
N-PSDB; ABK54194.
Identifying potential antibiotic or antimicrobial agent, comprises
contacting composition comprising pantothenate kinase (Coax) protein
with test compound and identifying inhibitor of the Coax protein -
Claim 10; Page 102-103; 128pp; English.
The invention describes assays for identifying a (potential) antibiotic
kinase (Coax) protein with a test compound, and determining the ability
of the test compound to inhibit the activity of the Coax protein, an
essential enzyme for the production of coenzyme A. Coax protein is a
valuable target for identifying bactericidal compounds. Coax modulating
agents can be used in an infectious animal model to determine the
efficacy, toxicity, or side effects of treatment with such an agent. This
is the amino acid sequence of a pantothenate kinase (Coax) protein
described in the invention.
Sequence 260 AA;
Query Match 43.5%; Score 575; DB 23; Length 260;
Best Local Similarity 44.1%; Pred. No. 1.5e-56;
Matches 113; Conservative 52; Mismatches 89; Indels 2; Gaps 1;
QY 1 MIFVLDVGNNTVILGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
DB 1 MLLAEQGNNTMFAIHGASWVAQWRSRSTRTADEYVWVLSQLSQGLGFRADAV 60
QY 61 IISVVPPIMPALERMCKLYPHIKPLIVGPGIKTGLDIKYDNPREGVADRVNVAAGIHL 120
DB 61 IISVVPPQSIENLRNLSRRYFNVEPLVIGENAKLIGDIVRIEKPSEAGADRLVNAIGAAMV 120
QY 121 YGSLPIIVDFGTATTTCYVINEHKQYMGGAIAPIGIMISTEALFARAALPRTIARPDII 178
DB 121 YPGPLVWDSGTATTFDVAADGAFEGGIIAPGINLSMQALHEAAAKLPRTIARPAQNR 180
QY 179 IIGKNTVSAMOGILYGVGVGVEGIVSRMKAKSKIPKPKVIATGGGLAPLIASESDIIDVVD 238
DB 181 IGVGDTVSAMOGVFWGIISLIEGLVARIKAEGERPEMTVIATGGVSLFEGATSDIDHFD 240
QY 239 PFLTITGLKLLYKENT 254
DB 241 SDLTITRGLLEIYRRNT 256
RESULT 13
AAU91153
ID AAU91153 standard; Protein: 258 AA.

THIS PAGE BLANK (USPTO)